**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A01N 43/04, 63/00, A61K 31/70, 48/00, C12N 1/00, 5/00, 5/06, 5/10, 15/00, 15/09, 15/11

A1

(11) International Publication Number: WO 96/22689

(43) International Publication Date: 1 August 1996 (01.08.96)

(21) International Application Number: PCT/US96/01337

(22) International Filing Date: 25 January 1996 (25.01.96)

(30) Priority Data:
08/378,235 25 January 1995 (25.01.95) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/378,235 (CIP)
Filed on 25 January 1995 (25.01.95)

(71) Applicant *(for all designated States except US)*: THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK [US/US]; West 116th Street and Broadway, New York, NY 10027 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PYLE, Anna, M. [US/US]; Apartment #18F, 560 Riverside Drive, New York, NY 10027 (US). MICHELS, William, J. [US/US]; Apartment #52, 664 163rd Street, New York, NY 10032 (US).

(74) Agent: WHITE, John, P.; Cooper & Dunham L.L.P., 1185 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AU, CA, JP, MX, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
*With international search report.*

(54) Title: MULTIPLE COMPONENT RNA CATALYSTS AND USES THEREOF

(57) Abstract

This invention is directed to a composition for catalyzed oligonucleotide cleavage comprising a synthetic non-naturally occurring oligonucleotide compound. The compound comprises nucleotides whose sequence defines a conserved group II intron catalytic region and nucleotides whose sequence is capable of hybridizing with a predetermined oligonucleotide target sequence to be cleaved, such target sequence not being present within the compound. The composition also includes an appropriate oligonucleotide co-factor. Preferably, the conserved group II intron catalytic region is a group II intron domain I catalytic region. In one embodiment the conserved group II intron domain I catalytic region may further comprise a conserved portion of a group II intron domain II, a group II intron domain III, a group II intron domain IV, a group II intron domain V, or a group II intron domain VI. The invention is also directed to methods of treatment and methods of use of such compounds.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| AM | Armenia | GB | United Kingdom | MW | Malawi |
| AT | Austria | GE | Georgia | MX | Mexico |
| AU | Australia | GN | Guinea | NE | Niger |
| BB | Barbados | GR | Greece | NL | Netherlands |
| BE | Belgium | HU | Hungary | NO | Norway |
| BF | Burkina Faso | IE | Ireland | NZ | New Zealand |
| BG | Bulgaria | IT | Italy | PL | Poland |
| BJ | Benin | JP | Japan | PT | Portugal |
| BR | Brazil | KE | Kenya | RO | Romania |
| BY | Belarus | KG | Kyrgyzstan | RU | Russian Federation |
| CA | Canada | KP | Democratic People's Republic of Korea | SD | Sudan |
| CF | Central African Republic | | | SE | Sweden |
| CG | Congo | KR | Republic of Korea | SG | Singapore |
| CH | Switzerland | KZ | Kazakhstan | SI | Slovenia |
| CI | Côte d'Ivoire | LI | Liechtenstein | SK | Slovakia |
| CM | Cameroon | LK | Sri Lanka | SN | Senegal |
| CN | China | LR | Liberia | SZ | Swaziland |
| CS | Czechoslovakia | LT | Lithuania | TD | Chad |
| CZ | Czech Republic | LU | Luxembourg | TG | Togo |
| DE | Germany | LV | Latvia | TJ | Tajikistan |
| DK | Denmark | MC | Monaco | TT | Trinidad and Tobago |
| EE | Estonia | MD | Republic of Moldova | UA | Ukraine |
| ES | Spain | MG | Madagascar | UG | Uganda |
| FI | Finland | ML | Mali | US | United States of America |
| FR | France | MN | Mongolia | UZ | Uzbekistan |
| GA | Gabon | MR | Mauritania | VN | Viet Nam |

# MULTIPLE COMPONENT RNA CATALYSTS AND USES THEREOF

This application is a continuation-in-part of U.S. Serial No. 08/378,235, filed January 25, 1995, the contents of which are hereby incorporated by reference into the present application.

## Background of the Invention

The invention disclosed herein was made with Government support under NIH Grants No. GM 50313-01, RO1 G50313, from the Department of Health and Human Services. Accordingly, the U.S. Government may have certain rights in this invention.

Throughout this application various references are referred to within parentheses. Disclosures of these publications in their entireties are hereby incorporated by reference into this application to more fully describe the state of the art to which this invention pertains. Full bibliographic citation for these references may be found at the end of this application, preceding the sequence listing and the claims.

In 1977 a discovery was made which radically modified our understanding of gene expression. Two laboratories observed that the protein coding sequences of an adenovirus gene are interrupted by regions of encoding DNA (Berget, 1977; Chow, 1977). The sequences of these interrupting regions, or "introns", are transcribed into precursor mRNA, and excised from the parent RNA in a process called RNA splicing. The coding ends of mRNA are then joined together, resulting in "processed" mRNA which is ready to code for protein. Comparison of genomic and RNA sequences revealed that introns are actually quite common in eukaryotic genes. Many genes contain multiple introns, and by splicing them out and religating the coding regions together in different combinations, a single gene can produce multiple protein products. This imparts an additional degree of control and

versatility to the expression of eukaryotic genes.

It was subsequently learned that a host of proteins and five nuclear RNAs (snRNAs) assemble as a complex upon nascent transcripts of pre-mRNA. This protein/RNA assembly was found to be responsible for catalyzing the splicing reaction and was appropriately dubbed the "spliceosome" (Brody, 1985). Analysis of reaction products showed that the spliceosome facilitates nucleophilic attack of an adenosine within the intron upon a phosphodiester bond at the 5'-intron/exon boundary to generate a branched intermediate (Fig. 1). The liberated 5'-exon then attacks at the 3'-exon-intron boundary, resulting in ligation of the two exons (thus generating coding mRNA) and excision of a lariat-shaped intron (Guthrie, 1991)(Fig. 1).

Not long after the discovery of RNA splicing, workers in the laboratory of Thomas Cech observed that precursor ribosomal RNA from Tetrahymena thermophila was capable of undergoing an in-vitro splicing reaction in the absence of protein enzymes (Kruger, 1982). Although it was assumed that a complement of endonucleases and ligases were involved in the two transesterifications required for splicing, the Cech lab showed that a structure within the folded intron itself catalyzed the splicing reaction in the presence of Mg2+ and a guanosine cofactor. This was the first example of a catalytic RNA, or ribozyme, in which RNA itself was capable of behaving like an enzyme. It appeared that "self-splicing" by Tetrahymena intron (phylogenetically classified as a group 1 intron) was fundamentally different from pre-mRNA splicing because of structural differences in their reaction products (Cech, 1986).

Several years later, members of the group II family of introns were observed to undergo a self-splicing reaction in-vitro (Peebles, 1986; van der Veen, 1986). Group II introns are unrelated to the group I introns in sequence or secondary structure and, although group II introns undergo

a self-splicing reaction, their reaction products are most similar to those of pre-mRNA splicing (Fig. 1). Like pre-mRNA splicing, the first step of group II self-splicing involves attack of an internal adenosine residue rather than a bound guanosine, and after splicing is complete, group II introns are excised as lariats. Additionally, the sequences surrounding the splice sites near group II introns and pre-mRNAs are very similar. For these reasons, group II introns have often been referred to as the "missing link" between group I introns, which are believed to represent a sort of molecular fossil of primitive RNA processing, and the splicing of pre-mRNA which is a relatively modern development (Cech, 1968; Jacquier, 1990). It has been proposed that the subdomains of the group II intron are analogous to snRNA molecules within the spliceosome and, since characterization of the unimolecular group II ribozyme would be far easier than dissection structure and mechanism of the spliceosome, study of group II intron ribozymes provides an important basis for our understanding of RNA splicing in general. Additionally, a functional analogy between group II intron subdomains and snRNAs would support the theory that eukaryotic pre-mRNA processing is catalzyed by RNA components of the spliceosome.

Instead of proteins and snRNAs, the active site of the group II intron is composed of a single strand of folded RNA. This RNA is divided into six domains (Fig. 2A-2B, Fig. 3), which may, like snRNAs and proteins, function separately in composing the active site and catalyzing the splicing reaction (Michel, 1989). It has therefore been of interest to define the functions of the individual domains. Techniques of molecular genetics have shown that Domains 2, 4 and most of 3 can be replaced by short hairpin loop regions without detectable effects on self-splicing (Bachl, 1990; Kwakman, 1989). Domain 1, which is the largest, recognizes the 5'-exon by base-pairing and participates in several other important stabilizing interactions (Jacquier,

1991). The most conserved region of the group II intron and the one with the greatest effects on splicing is Domain 5. This short hairpin region of RNA (only 34 nucleotides in the ai5g intron) appears to direct the folding or activity of the group II intron in a very fundamental way (Jarrell, 1988; Pyle, 1994).

In 1988, the laboratory of Philip Perlman found that a group II construct containing only Domains 1-3 was unable to undergo the first step of splicing. However, addition of the tiny Domain 5 in-trans rescued the splicing reaction (Jarrell, 1988). Direct proof of the strong D5 tertiary interaction was provided by kinetic and thermodynamic studies showing D5 bonds in trans with a $K_D$ of ~300nM. (Pyle, 1994). Given that Domain 5 is almost completely paired and there is no phylogenetic covariation between sequences in Domains 1 and 5 (Michel, 1989), it is highly unclear how it would associate with Domain 1 to generate the catalytically active structure.

With the finding that group II intron structures are formed during the trans-splicing of mitochondrial energetics genes from higher plants (Wissinger, 1992), it is now believed that group II splicing reactions are of general importance. They are central to the metabolism of plants, yeast and possible other eukaryotes. Some group II introns contain open reading frames which code for "maturases', or proteins which assist in the splicing and reverse-splicing of group II introns in-vivo (Perlman, 1989). Some of these maturases contain reverse transcriptase and RNase H domains (Lambowitz, 1990). These findings have precipitated interest in the introns as mobile genetic elements, or "infectious" RNAs, readily able to integrate into new genomes (Lambowitz, 1989). Interestingly, open reading frames and base-pairing recognition domains for trans-splicing variants of the group II intron are usually found in Domains 2 or 4 of the group II structure, regions of high variability and low sequence conservation (Michel, 1989).

Although group II intron reactions are important in their own right, the greatest interest in group II introns stems from the hypothesis that they are related chemically and structurally to the spliceosomal apparatus responsible for eukaryotic pre-mRNA processing (Sharp, 1985). This theory implies that the group II intron may be used as a model for reactions which take place in the spliceosome, a reaction center which is extremely difficult to characterize because it contains hundreds of components and must be assembled in cell extracts. It also suggests that group II intron chemistry will describe "remnants" of RNA catalysis still present in higher biological systems and suggests that spliceosomal processing is RNA-catalyzed despite the presence of proteins. There is a great deal of chemical theory which now supports these ideas. Work by Moore, Sharp and coworkers on the stereochemistry at pre-mRNA splice sites has provided evidence for the involvement of two active sites in spliceosomal catalysis (Moore, 1993). This suggests that group II introns may also have two active sites; one for the first step and one for the second step of splicing. Group I introns are generally believed to have only one active site, and that is why the stereochemistry of nucleophilic attack in the first step is opposite to that of the second step (McSwiggen, 1989; Rajogopal, 1989). Spliceosomal processing and group I introns are most similar in the second step of splicing, where both prefer the pro-S isomer of phosphate. Spliceosomal processing and group II introns appear most similar in the first step which is initiated by adenosine 2'-OH attack. This implies that the first step in group II intron splicing is most likely to represent a new form of catalytic active site and one which we must characterize before we can understand the relationship to pre-mRNA processing. To this end, engineering multiple-turnover ribozymes from the group II intron will enable us to examine the chemical mechanism for reactions catalyzed by the active site(s) of the group II intron. This same approach was a prerequisite to

characterization of the transition state, the active-site structure and the detailed mechanism descriptive of the group I intron (Cech, 1992).

It was originally believed that in-vitro group II self-splicing was inefficient before appropriate reaction conditions were defined in 1988 (Jarrell, 1988). Despite splicing reactions which are complete within minutes and the fact that splicing is a complicated multi-step process, the belief that group II introns have lumbering reactivity still persists. It is also commonly believed that the structure of the group II intron is metastable, with few molecules in a pool assuming an active conformation. Unfortunately, there is little quantitative data to substantiate either of these theories. Many generalized ideas about the nature of the group II intron come from observations of a single molecule, particularly the ai5g intron from yeast mitochondria. There have been no systematic comparisons of group II intron reactivity of stability, and virtually no quantitative analysis on the kinetic behavior of any single intron. We know nothing about the architecture of the folded structure. Clearly, before the group II intron can be discussed coherently as a model for pre-mRNA processing, we must characterize much of its fundamental behavior. As indicated above, the most rapid way to do this is to transform the unimolecular group II self-splicing reaction into multiple-turnover ribozyme that function under simple conditions. This will impart significant control over the reaction and facilitate careful mechanistic analysis of the intron.

Several research groups have been pursuing the use of catalytic RNA, also known as ribozymes, for uses in diagnostics, as restriction enzymes, see for example Cech et al. U.S. Patent No. 4,987,071. Ribozymes which are highly specific in the recognition sequences offer potential in therapy for plants or mammals, see for example Haseloff et

al. U.S. Patent No. 5,254,678 for hammerhead ribozymes, Jennings et al. U.S. Patent No. 5,298,612 for minizymes; Hampel et al. European Appn. No. 89 117 424.5, filed September 20, 1989 for hairpin ribozymes; and Altman et al. U.S. Patent No. 5,168,053, for prokaryotic RNAase P. However, this is the first report of the use of a modified group II intron composition to act in trans, i.e. as a catalyst, to cleave a nucleic acid target.

## Summary of the Invention

This invention is directed to a composition for catalyzed oligonucleotide cleavage comprising a synthetic non-naturally occurring oligonucleotide compound. The compound comprises nucleotides whose sequence defines a conserved group II intron catalytic region and nucleotides whose sequence is capable of hybridizing with a predetermined oligonucleotide target sequence to be cleaved, such target sequence not being present within the compound. The composition also includes an appropriate oligonucleotide co-factor. Preferably, the conserved group II intron catalytic region is a group II intron domain I catalytic region. In one embodiment the conserved group II intron domain I catalytic region may further comprise a conserved portion of a group II intron domain II, a group II intron domain III, a group II intron domain IV, a group II intron domain V, or a group II intron domain VI. The invention is also directed to methods of treatment and methods of use of such compounds.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Shows a comparison of RNA processing by self-splicing introns and the spliceosomal apparatus.

Figure 2A and 2B: Shows the secondary structure of the two group II introns from yeast mitochondria (seq ID No. 4 and 5).

Figure 3: Shows a schematic of a group II intron. The structure is composed of six subdomains organized like spokes of a wheel. Domain 1 contains the EBS1 and EBS2 regions (Exon-Binding Sites 1 and 2) which base pair to the IBS1 and IBS2 segments of the intron. The first step of splicing is initiated by attack at the 5'-splice site from an adenosine 2'-OH group on Domain 6. Domains 2-4 are dotted because they are dispensable for this first step.

Figure 4A and 4B: Cleavage of a short RNA substrate (S) by the group II ribozyme. (Fig. 4) Domain 1 (D1) combines with Domain 5 (D5) and substrate (S) to catalyze specific cleavage at a site analogous to the 5'-splice site. Shaded regions and dashes between D1 and S designate pairing interactions IBS1:EBS1 (closest to the cleavage site) and IBS2:EBS2. D1 RNA (425 nts) was transcribed from plasmid pT7D1 cut with Eco RI. This plasmid was constructed by PCR amplification of D1 from the full-length plasmid pJD20 (Jarrell, 1988). Primers specifically amplified a 425 nt intronic fragment including all of D1 beginning with the +1 nucleotide of the intron (G). The upstream primer included a T7 promoter sequence and a BamHI restriction site. The downstream primer contained an EcoR1 site. The PCR fragment was ligated into a pUC19 vector, expressed and sequenced. D5 was transcribed as previously described from plasmid pJDI5'-75 (kindly provided, together with pJD20 by Dr. Philip S. Perlman) (Pyle, 1994). Transcribed substrates were made using T7 RNA polymerase using synthetic DNA oligonucleotide templates according to published procedures.

All RNAs were purified as described (Pyle, 1994). (Fig. 4) Ribozyme cleavage of substrate (S) 5'-$^{32}$p-GGAGUGGUGGGACAUUUUC˅GAGCGGUU-3' (SEQ ID NO 1) to a 5'-end-labelled 19 nt. product (p) and an 8 nt. unlabelled product. The cleavage site is marked with a "˅" and IBS sequences are underlined. Lanes 1-8 correspond to a reaction of S(1 nM) in the presence of saturating D1 (100 nM) and saturating D5 (3 μM). The reaction was performed under standard reaction conditions: at 42°C in 80 nM MOPS (pH 7.5), 1M KCl, and 100 nM $MgCl_2$. RNAs were mixed at 1.67x final conc. in 80 nM MOPS pH7.5 and then heated to 95°C for 1 min. to denature potentially mis-folded RNAs formed during storage. After cooling to 42°C, reaction was initiated by addition of 100 nM $MgCl_2$, 1M KCl and 80 nM MOPS (final concentrations). Aliquots of the reaction were removed at the time intervals specified in the figure, combined with denaturing dye on ice, loaded on a 20% denaturing polyacrylamide gel (PAGE) and quantitated as previously described.

**Figure 5A and 5B: Single-turnover kinetics of S cleavage by the three-part group II ribozyme.** Rate constants and kinetic parameters (Table 1) were obtained by plotting pseudo-first order rates ($k_{obs}$), determined from the slope of inset plots against a reactant varied in concentration. It was determined that $k_{obs}$ values at saturation did not vary from 0.01 to 3 Nm [S], indicating that pseudo-first order kinetics apply (data not shown). Pseudo-first order semi-log plots of the cleavage timecourse were linear for >3 reaction half-times (see also Fig. 9B). Early time points (generally <50% completion) were used in the calculation of $k_{obs}$ (see also Fig. 9B). Individual values of $k_{obs}$ were determined over a 6 week time period with the same RNA stocks and varied by 18% (with 95% confidence and 5 trials). The 24-nucleotide substrate, CGUGGUGGGACAUUUUC˅GAGCGGU (SEQ ID NO 2), was made synthetically (Scaringe, 1990). The empirically determined endpoint for reaction of this substrate was 80% after >7 reaction half-times. This value

was used in correction of all pseudo-first order plots used in this figure. (Fig. 5A) Inset: Representative pseudo-first order plots for cleavage of S (0.05 nM) at saturating D5 (3 μM) and varying concentrations of D1 (0.5, 1.5, 3.25, 5, 10, 20, 100 nM shown). Hyperbolic plot: The $k_{obs}$ values determined from inset plots were graphed against [D1] to generate the apparent binding curve for D1. In both A and B hyperbolic plots, the $k_{max}$ value for the reaction is the horizontal asymptote of plots (~ 0.02 $min^{-1}$), and Km is the concentration of enzymatic subunit at which $k_{obs}$ is one-half $k_{max}$. (Fig. 5B) Inset: Representative pseudo-first order plots for cleavage of S (0.1 nM) at saturating D1 (25 nM) and varying amounts of D5 (0.2, 0.33, 0.67, 1, 3, 9 μ.M shown). Hyperbolic plot: The $k_{obs}$ values determined from inset plots were graphed against [D5] to generate the apparent binding curve for D5 and the kinetic parameters shown in Table 1 (line 3).

**Figure 6: Ribozyme cleavage of a single phosphorothioate linkage:** Comparison of ribozyme cleavage of a transcribed all-phosphate RNA substrate (squares: connected by ---) and a synthetic substrate containing a single phosphorothioate at the cleavage site (circles; connected by - - -). A standard ribozyme digestion was performed as in Figures 4 and 5, except that, unlike these previous kinetic figures, final [KCl] was 2M in order to approach endpoint more rapidly. Doubling the KCl concentration increases the rate of reaction approximately ten-fold, while pseudo-first order plots remain linear (data not shown). The all-phosphate substrate was cleaved to 96% ± 0.8% (with 95% confidence and 4 trials), empirically determined after >7 half-times. This substrate was cleaved 52% ± 4% (with 95% confidence and 5 trials) at >7 half-times (see Fig. 7, lane "M"). Time-courses using newly synthesized RNA reproducibly resulted in endpoints slightly less than 50%. Phosphorothioate linkages slowly convert to phosphates during storage and this factor may explain why slightly more than 50% of the substrate is

cleaved (Herschlag, 1991). Time-courses using this oligonucleotide after it was first synthesized were cleaved to <48%. Including 10 mM DTT in the reaction had no effect on the extent of cleavage.

**Figure 7: Determining the stereochemical identity of the less reactive substrate.** Three different synthetic substrates (0.4 nM) were digested with 3 U/μl of snake venom phosphodiesterase (SVPD, Boehringer-Mannheim), in a reaction containing 0.1μg/μl carrier tRNA, 50 mM Tris-HCl pH 8.5 and 10 mM $MgCl_2$. Incubations were at 37°C for 0, 1, 2, 5, 10 and 30 minutes, respectively. The lanes marked "phosphate" represent digestion of a normal all-phosphate substrate (described in Fig. 5). The lanes marked "Rp + Sp" are digestions of a synthetic substrate containing a single phosphorothioate at the cleavage site (described in Fig. 6). The lanes marked "Sp" are digestions of the synthetic phosphorothioate substrate that was less reactive in the presence of the ribozyme. To isolate this substrate, the racemate (Rp+Sp) was cleaved to apparent completion by the ribozyme (as in Fig. 9), resulting in the 50% cleavage shown in lane 'M". The less reactive band was re-isolated by PAGE and purified in the same manner as the other two substrates before treatment with SVPD. Radioactive counts were quantitated and the amount of stalling at +1G was expressed as a fraction of the whole lane. This stall product fraction was integrated over the timecourse of the reaction to give relative amounts of exonucleolytic stalling between the three time courses. The two reactions (Rp+Sp, and Sp) that had not reached full digestion by 30 min. were extrapolated to theoretical endpoints using the slope of the last two time points. Both substrates gave extrapolated endpoints of 42 min. The final ratio of stall product at +1G was {0.05 : 0.59 : 1} for {phosphate : Rp+Sp : Sp}, respectively.

**Figure 8A and 8B: Nuclease mapping of transcribed**

**substrates and their cleavage products.** In Fig. 8A: nuclease digestions of 5'-$^{32}$P end-labelled all-phosphate substrate described in the legend to Fig. 4. In Fig. 8B, the 5'-$^{32}$P end-labelled substrate was transcribed from the same DNA template, but α-thio (Sp) GTP (Amersham) was used as the sole source of GTP in the reaction. This substrate contains an Rp phosphorothioate residue at every G, including the cleavage site. Enzymatic digestions were carried out under standard conditions (Donis-Keller, 1977). For both A and B, Lane 1 is the full-length substrate. Lane 2 and 3 are digestions with hydroxide. Lane 4: Digestion with endonuclease U2 (BRL, cuts ApN). Lane 5: enzyme B. Cereus (Pharmacia, cuts UpN or CpN), Lane 6: enzyme CL3 (BRL, cuts CpN), Lane 7: enzyme T2 (Pharmacia, cuts GpN), Lane 8: Ribozyme reaction product (19-mer RNAs), Lane 9: P1 nuclease (Boehringer Mannheim, no sequence specificity), Lane 10: Hydroxide ladder on substrate, Lane 11: P1 digestion of ribozyme reaction product (all-phosphate substrate only). Sequence specific nucleases cleave RNA to leave terminal 3' phosphate, while alkaline hydrolysis leaves predominantly 2'-3'-cyclic phosphates. P1 nuclease cleaves RNA to leave 3'-OH.

**Figure 9A and 9B: (Fig. 9A). Quantitative determination of (Rp) phosphorothioate cleavage by the ribozyme.** Phosphorothioate substrate was made by transcription of DNA template as in figure 8B, except labelling was accomplished by including a $^{32}$P CTP in the transcription reaction. Nuclease digestion of precursor substrates: RNA (3000 cpm) was incubated with 0.01 U/μl P1 nuclease in 25 mM acetate buffer pH 5.3, 10 mM DTT, and 0.05 μg/μl tRNA competitor for 30 min. at 37°C. Lane 1. Digestion of all-phosphate substrate with P1 nuclease. Lane 2: All-phosphate substrate digested with ribozyme, followed by P1. Lane 3: Digestion of thio-substrate with P1. Lane 4: Digestion of thio-substrate with ribozyme, followed by P1. Lanes 5 and 6: Digestion of thio-substrate with D1 and D5 alone,

respectively, followed by P1. All ribozyme incubations were performed to completion under the conditions described in Figure 6. following ribozyme digestion, RNAs were ethanol precipitated, resuspended and digested to completion with P1 nuclease (Lanes 1+3: 0.01 U/$\mu$l P1 nuclease in 25 mM acetate buffer pH 5.3, 10 mM DTT and 0.05 $\mu$g/$\mu$l tRNA competitor for 30 min. at 37°C. Lanes 2, 4, 5 and 6: as in Lanes 1 and 3 except tRNA was 0.025 $\mu$g/$\mu$l). Any radioactivity in the middle (*pCp$_s$G) band in lane 4 corresponds to counts from the small fraction of thio-substrate (~ 18%) that was uncleavable by the ribozyme. Note that there cannot be a C 3' to the cleavage site, that is G+1 cannot be a C. (Fig. 9B). **Kinetics of transcribed ribozyme substrates:** Cleavage of the (Rp)-phosphorothioate substrate (circles, described in Fig. 8 and 9A) and the transcribed all-phosphate substrate (squares) follow pseudo-first order kinetics for more than three half-lives. Reaction conditions and saturating RNA concentrations were identical to Figure 4, except [S] concentration was 0.1 nM. Values for $k_{obs}$ are 0.00094 and 0.032 $min^{-1}$, respectively (average of two determinations). For the thio-substrate, data is corrected for an empirically determined end-point (at 7 half-times) of 83%. The all-phosphate substrate data is corrected for an empirically determined endpoint of 96% (at >7 half-times).

**Figure 10A-10B: Secondary structures of domains used in investigating the first step of group II intron splicing.** (A) Schematic secondary structure of a group II intron and the two transesterifications involved in self-splicing. During the first step of splicing (labeled step 1), a 2'-hydroxyl group on the indicated adenosine of D6 attacks the 5'-splice site (shown as an open circle). During the second step of splicing (labeled step 2), the free 3'-hydroxyl of the 5'-exon attacks the 3'-splice site (shown as a closed circle). Regions of Domain 1 (EBS1 and 2) that base-pair with the 5'-exon (IBS1 and 2) are shown as wide lines (Jacquier and Michel, 1987; Jacquier and Michel, 1990).

Parts of the RNA that were not present in this study are shown as dotted lines, while RNA that was present is shown as a solid line, including 5'-exon/Domains 1-3 (exD123, from the 5'-end of the RNA to the double bars between Domains 3 and 4), and Domain 5-6 (D56, the region between double bars flanking Domains 5 and 6). This schematic does not attempt to illustrate the exact secondary structure of individual domains, which are much more complex than indicated here (Michel et al., 1989). (B) The nucleotide sequence of D56 (91 nts). The line at the 5'-end represents 15 nucleotides of vector sequence left over from cloning. The 3'-end of D56 terminates at the last nucleotide of the intron. 3'-exon nucleotides have been removed so that the second transesterification of splicing cannot occur.

**Figure 11A, 11B and 11C: Competition of hydrolysis and transesterification during the first step of splicing. (A).** Time courses are shown for reaction of internally-labeled $^{32}$P-exD123 (1 nM) with excess D56 or D56A- (3 $\mu$M) at 45¡C using the standard reaction buffer at two different pH values. The indicated bands are precursor (exD123, 1003 nt), D56/D123 (D56 that is connected to D123 through a 2'-5' branch, 801 nt), D123 (linear intron resulting from hydrolytic 5'SS cleavage, 710 nt), D12 (a secondary product of hydrolysis, 560 nt), and 5'-exon (293 nt). The D56/D123 band occurs in the presence of D56 but is not found in the presence of D56A- (see +D56 lanes). However, hydrolysis products are observed in reactions of both D56 and D56A- (all lanes except t = 0). Hydrolysis reactions dominate over the branching reaction at long reaction times, especially at higher pH (+D56 lanes, pH 7.5). (B and C) Detailed time-courses of the reaction shown in (A), between $^{32}$P-exD123 and D56 at pH 7.5. The evolution of hydrolysis product (B) simultaneous with branched product (C) shows different kinetic behavior. The branch bursts quickly and then declines while hydrolysis products arise in pseudo-first order fashion, fitting an exponential curve.

**Figure 12A, 12B and 12C. Timecourse of branching independent of hydrolysis and characterization of branched product.** (A) Autoradiogram of $^{32}$P-D56 (1 nM) converting to $^{32}$P-D56/D123 in the presence of exD123 (2 $\mu$M) under standard reaction conditions at pH 7.0. At longer times, bands develop below $^{32}$P-D56 and $^{32}$P-D56/D123. These are degradation products of both species that accumulate over time and do not affect kinetic analysis of the reaction. (B) Plot of fraction reacted versus time, quantitated from the data shown in (A). The reaction shows exponential behavior that levels off at 48% total reaction after several hours. (C) Digestion of $^{32}$P-D56 (left) and $^{32}$P-D56/D123 (right) using P1 nuclease, which cannot cleave branched trinucleotides. The $^{32}$P-D56 is internally labeled with $\alpha$-$^{32}$P-UTP, so P1 digestion will result in a radioactive trinucleotide only if branching occurs specifically to produce pA(2'-pG)3'-pU (Reilly et al., 1989). The trinucleotide was resolved from mononucleotide on a one-dimensional 20% polyacrylamide gel (Reilly et al., 1989; Moore and Sharp, 1993). As expected, only P1 treatment of the branched species (right) resulted in $^{32}$P-labeled pA(2'-pG)3'-pU, while P1 digestion of $^{32}$P-D56 results only in mononucleotide (left). RNAs were reacted with P1 for 12.5 minutes in the experiment shown, although the same results were obtained from 5-15 minutes. The ratio of branched trinucleotide to mononucleotide was 1:54, whereas the expected ratio was 1:21 given the uridine composition of D56. This higher ratio may be due to hydrolytic instability of the branch during workup or occasional branching from the wrong nucleotide. The latter explanation is unlikely given that D56A- does not branch (Figure 11A).

**Figure 13A,13B: Single-turnover kinetic analysis of the forward branching reaction results in a $K_s$ curve for D56.** Figure 13A. Reactions such as those shown in Figure 12A were performed with increasing concentrations of exD123. The

resultant $k_{obs}$ values (determined as shown in inset) were plotted as a function of [exD123] and an apparent binding curve is generated, reflecting the affinity of D56 for exD123 . Using a graphing program, the data were fit to a standard equation describing 1:1 bimolecular association, resulting in a $K_m$ of 510 ± 160 nM and $k_{cat}$ of 0.041 ± 0.004 $min^{-1}$ (Pyle and Green, 1994). Standard errors shown were obtained from these fits as described previously. All semilog plots used for determination of $k_{obs}$ values were corrected for a 51% endpoint determined empirically after 7 half-times. Variance in individual $k_{obs}$ values, the points on this curve, is described below. (Fig 13B): Semilog plot of $^{32}$P-D56 (1 nM) reaction with exD123 (2 μM), having corrected the reaction for a 48% endpoint (shown in Figure 12B). The observed rate of reaction is obtained by solving the equation of the line for the reaction half-time, as described in the experimental section. The $k_{obs}$ for the inset plot shown is 0.033 $min^{-1}$. A small burst of initial rate (at 4% reaction) can be seen by the fact that a line reproducibly crosses the y-axis below zero. Individual $k_{obs}$ values obtained on different days, with different RNA stocks, were found to have a variance of ±37% with 95% confidence based on six measurements. This is a very conservative estimate of maximum error given that the variance in timecourses performed on the same day is observed to be much lower. For example, the D56 and D56A-hydrolysis reactions, studied side by side on the same day, both yielded $k_{obs}$ values of 0.018 $min^{-1}$ (see text).

**Figure 14A, 14B, 14C : Kinetic analysis and characterization of the reverse-branching reaction: (A)** Timecourse for the reverse branching reaction: Upon addition of saturating 5'-exon, the $^{32}$P-D56/D123 branched species (1 nM) converts to $^{32}$P-D56 under standard reaction conditions at pH 7.0 (lanes 1-10, reaction timepoints of 0.75, 1.0, 1.5, 3.5, 7.5, 10, 15, 30, 60 and 120 minutes). In the absence of 5'-exon, 32P-D56 is liberated very slowly,

as a product of hydrolytic attack on the 2'-5' link (lanes 11-18, reaction timepoints of 0, 2.5, 5, 10, 15, 30, 60, and 120 minutes). By quantitating the amount of random breakage in the branched molecule and comparing it to hydrolysis of the 2'-5' linkage, we determined that the 2'-5'-linkage is approximately 130-fold more labile than surrounding 3'-5' linkages in this context (lane 18). (B). Covalent attachment of 5'-exon to D123 in the reverse branching reaction: The 5'-exon (2 nM) was $^{32}$P-end labeled and reacted with D56/D123 (0.1 nM, contained a trace of radioactivity for purposes of isolation, lanes 3-5) and $^{32}$P-D56/D123 (0.4 nM, lanes 6-8) under standard reaction conditions at pH 7.0. Timepoints were 30 (lanes 3 and 6), 60 (lanes 4 and 7) and 120 minutes (lanes 5 and 8). Lane 1 is $^{32}$P-D56; lane 2 is $^{32}$P-5'-exon (-17/0); and lane 9 is $^{32}$P-D56/D123. In lanes 3-5, $^{32}$P-5'-exon reacts and shifts up to form band $^{32}$P-ex/D123. In lanes 6-8, the same shift is observed, concomitant with generation of equivalent molar amounts of free $^{32}$P-D56 (there are two-fold more counts in the $^{32}$P-ex/D123 band because it is end-labeled while D56 is body-labeled). (C) Kinetic analysis of the reverse reaction: The data shown in (A) were plotted, including additional early reaction timepoints. The fraction reacted vs. time fit a biexponential equation representing a combination of two rates (experimental section). The first phase has a rate of 0.95 ± 0.17 $min^{-1}$. A total of 19 ± 1.3% of the precursor reacts through this phase. The second phase proceeds with a rate of 0.011 ± 0.0010 $min^{-1}$. The variances shown are the standard errors calculated from fit to the equation described in Experimental Details below.

**Figure 15: Overall reaction mechanism and kinetic parameters for the first step of group II intron self-splicing.** D56 binds to exD123 with a $K_d$ of 0.51 ± 0.16 μM (Figure 13). There are three reactions that begin to compete after complex formation occurs: Whether or not it is in the proper conformation, the D56/exD123 complex can

catalyze hydrolysis at the 5'-splice site with a chemically limited rate of 0.018 ± 0.0070 $min^{-1}$ ($k_{chem}$, Figure 11 and text). At the same time, the D56/exD123 complex samples conformational states, eventually reaching one that is capable of branching (schematized as a bend in D56). This process occurs with a rate of 0.041 ± 0.0040 $min^{-1}$ ($k_{f(conf)}$, Figure 13). At this point, the 5'-exon can dissociate ($K_d$ ~ 10 nM) or it can attack the branch site. The branched species is more conformationally constrained, and 19% of it reacts rapidly with 5'-exon at a rate of 0.95± 0.17 $min^{-1}$ ($k_{r(burst)}$, Figure 14C). The remaining molecules sample an active conformation and react with a rate of 0.011 ± 0.001 $min^{-1}$ ($k_{r(conf)}$, Figure 14C).

**Figure 16A-16B: Catalytic mechanism of RNA cleavage by the minimal group II ribozyme and schematic representation of substrate recognition and cleavage.** Figure 16A. The catalytic mechanism of RNA cleavage is shown; water (or hydroxide) is the nucleophile. Figure 16B. The minimal group II ribozyme comprises domain 1 (D1, nts=nucleotides) of the six domains present in the full-length intron and contains two exon-binding sites (EBS1 and 2). The EBSs hybridize to intron binding sites (IBS1 and 2) on the oligonucleotide substrate, separated by a three nucleotide link. In the presence of a second RNA moleucle containing domain 5 (D5) of the group II intron, the substrate is cleaved at the site indicated by the arrow.

**Figure 17: Determination of pseudo-first order rate constants for the cleavage of chimeric substrates at ribozyme saturation.** Rates were measured simultaneously, under standard reaction conditions where the concentration of D1 was 100 nM, the concentration of D5 was 3 μM (experimental section). Rates of reaction ($k_{obs}$) were determined by correcting each data set for the amount of total inactive substrate and plotting the data as shown, where P = fraction product at each timepoint. By solving

the equation for each line, the half-time of reaction and subsequently the $k_{obs}$ were determined as described previously [Michels and Pyle, 1995]. For the the rapidly reacting substrates shown, the rates were 0.021 $min^{-1}$ for rS, 0.018 $min^{-1}$ for -3dU, 0.026 $min^{-1}$ for -4dU and -5dU, 0.025 $min^{-1}$ for -6dA and 0.020 $min^{-1}$ for -7dC. The variance in these rates was ±16% determined using eight separate trials. For -1dC the reaction rate was 0.0013 $min^{-1}$, with a variance of ±23% determined using eight different trials. For -2dU, the rate was 0.0042 $min^{-1}$, with a variance of ± 29% based on three trials. All errors were calculated using a 95% confidence limit and trials were performed on different days. For experiments performed side-by-side as shown here, relative error is expected to be much lower than the variances conservatively stated above.

**Figure 18: Mapping the position and termini of substrate cleavage products:** High-resolution mapping with endonucleases and alkali: Lanes 1 to 7 represent mapping of $^{32}P$ 5'-end labeled rS, lanes 8 to 14 are maps of $^{32}P$ 5'-end labeled -1dC. For each substrate, lanes 1 and 14 are full-length substrate; lanes 2 and 13 are partial digestions of substrates by T1 nuclease (Pharmacia, cuts GpN); lanes 3 and 12, are partial digestions with enzyme *B. Cereus* (Pharmacia, cuts UpN or CpN); lanes 4 and 11 are partial hydroxide digestions of the full length substrates; lanes 5 and 10 are partial digests on the 50/50 mix of precursor and product from a ribozyme reaction; lanes 6 and 9 are the 50/50 mix of product and precursor from a partial ribozyme reaction; lanes 7 and 8 are partial digestions with nuclease P1 (Boehringer Mannheim, cuts DNA or RNA). The mobilities for substrate (S) and product (P) are marked on the side of the gel. Bands in lanes 6, 7, 8 and 9 have a 3'-hydroxyl terminus, causing them to migrate more slowly than their corresponding bands in other lanes, which possess a 3'-phosphate or 2'-3'-cyclic phosphate terminus. There are also mobility differences due to the presence of a single

deoxynucleotide, which tends to make chimeric oligonucleotides migrate more rapidly (lanes 6-9; lanes 7 and 8). Verification of deoxynucleotide cleavage by characterizing the 3′ termini of reaction products. A 50/50 mixture of oligonucleotide substrate and product from a partial D1/D5 ribozyme reaction was subjected to oxidation (+ $NaIO_4$) and then 3′-end labeled with a hydrazine dye derivative (+ Lucifer Yellow). Lanes 1 to 4 are rS substrate, and lanes 5 to 8 are -1dC substrate. The identity of each band is indicated to the right. After oxidation and covalent dye-labeling, both the precursor and product band of the rS substrate undergo a large mobility shift (lane 4). For -1dC, only the substrate band shifts to slower mobility (lane 5), indicating that the cleavage product of -1dC has a 3′ deoxy terminus, which cannot be oxidized with $NaIO_4$. Doublets in lanes 4 and 5 are most likely due to the products of Lucifer Yellow hydrazide attack on one or both of the aldehydes produced by oxidative attack of the ribose sugar [Hannske et al., 1974]. Following oxidation with $NaIO_4$, there is a small mobility shift for species with a 3′-ribose because the sugar is oxidized to a larger dialdehyde moiety (compare lanes 1 and 2). Additional bands in the +$NaIO_4$ lanes are attributed to periodate attack on residues other than the ribose sugar [Rizzino and Freundlich, 1975]. Verification of deoxynucleotide cleavage by the D123 ribozyme. This experiment and figure is the same as in (B), except that D123 was the ribozyme instead of D1. As in (B), the -1dC cleavage product is shown to have a 3′-deoxynucleotide terminus by inspection of lane 5.

**Figure 19A, 19B and 19C: Determination of kinetic parameters for rS, -1dC and -2dU substrates:** By plotting $k_{obs}$ as a function of [D1] at saturating [D5], one can fit the data to a binding curve and obtain kinetic parameters for a reaction that follows Michaelis-Menten kinetics. Data were fit as previously described [Michels and Pyle, 1995; and Pyle and

Green, 1994]. (A) For the rS substrate, $K_m$ = 4.9 nM and $k_{cat}$ = 0.025 $min^{-1}$. (B) For -1dC, $K_m$ = 14 nM and $k_{cat}$ = 0.0019 $min^{-1}$. (C) For -2dU, $K_m$ = 10 nM and $k_{cat}$ = 0.0051 $min^{-1}$. All $k_{cat}$ values determined from these fits are in good agreement with the saturating $k_{obs}$ values reported in Fig. 1. Errors and $k_{cat}/K_m$ values determined from this experiment are provided in Table 2.

**Figure 20: Determination of saturating $k_{obs}$ values for oligonucleotide cleavage by the fast D123 ribozyme.** Cleavage of rS and -1dC substrates by D123 (100 nM) and D5 (3 $\mu$M). The apparent $k_{obs}$ values from this data are 0.88 ± 0.089 $min^{-1}$ for cleavage of rS and 0.17 ± 0.021 $min^{-1}$ for -1dC. Errors were calculated with 95% confidence from 3 trials, using average $k_{obs}$ values of 0.84 $min^{-1}$ and 0.16 $min^{-1}$ for rS and -1dC, respectively. Because D123 cleaves oligonucleotides so efficiently, rate-limiting conformational processes become evident in the seconds timescale. A small initial lag exists at timepoints earlier than 15 seconds for both substrates. By conducting fluorescence binding experiments and altering the order of component addition, the lag was found to derive from a folding event related to the association of D5 (Z. F. Qin, data not shown). After this initial lag, the data are almost perfectly linear such that, at longer reaction times, the folding event is not rate-limiting. Rates were calculated from the fit to the linear portion of this data (after 15 seconds).

**Figure 21: Summary of deoxynucleotide effects on the chemical rate of D1/D5 ribozyme cleavage:** The magnitude of $k_{chem}$ diminution is plotted for the single and block deoxynucleotide substitutions along the oligonucleotide substrate strand. The $k_{chem}$ effect for IBS2 is labelled as WT for block because the $k_{cat}$ for cleavage of the d(IBS2) oligonucleotide was the same as the rS substrate, provided that additional D1 was added to overcome slight weakness in

the chimeric EBS2/IBS2 helix.

**Figure 22A, 22B and 22C:** The three reactions observed during in-vitro self-splicing of group II intron ai5g. The reaction products shown at right are the species observed on high-resolution polyacrylamide gels of self-splicing reaction products. The second step of splicing is extremely rapid and the first step of splicing is rate limiting. Therefore, although splicing intermediates are shown in (A) and (B), these are barely detectable, if they are observed at all. SER(C) can take place immediately after splicing, so its products (free exons) are also evident on separation gels.

**Figure 23A, 23B and 23C:** High-resolution sequencing gel of reaction products from self-splicing. The product band are labeled as follows: A is lariat intron; B is precursor RNA; C is linear intron (which comigrates with broken lariat); D is spliced exons; E is free 3'-exon; F is free 5'exon. In all cases, linger timecourses than those shown were required for kinetic analysis (A) Splicing in 100 mM $MgCl_2$ and 500 mM $(NH_4)_2SO_4$: Lanes 1-15 correspond to timepoints of 0,1,2,3,4,8,10,15,20,30,40,60,120,240, and 900 minutes, respectively. (B) Splicing in 100 mM $MgCl_2$ and 500 mM KCl: Lanes 1-6 correspond to timepoints of 0,10,15,20,30, and 40 minutes, respectively. (C) Splicing in 100 mM $MgCl_2$ and 500 mM $NH_4Cl$: Lanes 1-5 correspond to timepoints of 12,20,30,45, and 60 minutes, respectively. Splicing products D-F are also observed under the latter two sets of conditions, although the lower half of the gels are omitted. These shorter products were identical to those shown in full lanes of Figure 8.

**Figure 24A, 24B, 24C and 24D:** Timecoureses for self-splicing under low-salt conditions. Experimental data for the decrease in precursor RNA (□), and the evolution of linear (●) and lariat (▲) intron products are plotted and fit

using equations 1-3 in the text. Both reactions proceed to completion at very long times, but the significant amount of RNA degradation that occurs after five hours at pH 7.5 precludes quantitative analysis after that time. (A) Reaction profile in 100 mM $MgCl_2$. The $k_T$ value determined from decrease in precursor was 0.0049±0.00022 $min^{-1}$. The coefficients $f_{hy}$ and $f_{br}$, indicating the relative amounts of linear and lariat intron were 0.75±0.11 and 0.18±0.0060, respectively. These represent the horizontal asymptotes of the product curves shown on the plot. The $k_T$ estimations determined from the fit of linear and lariat evolution were 0.0047 $min^{-1}$ and 0.0096 $min^{-1}$, respectively. Because they are based on a relatively smaller number or radioactive counts, these values are less accurate than $k_T$ determined from decrease in precursor. (B) Reaction profile in 10 mM $MgCl_2$ and 5 nM protamine. The $k_T$ value determined from decrease in precursor was 0.0029±0.00014 $min^{-1}$ and the coefficients $f_{hy}$ and $f_{br}$ were 0.77±0.088 and 0.14±0.020, respectively. The $k_T$ estimations determined from the fit of linear and lariat evolution were 0.0030 $min^{-1}$ and 0.0056 $min^{-1}$, respectively. (C and D) Kinetic parameters of ribozymes at low salt conditions.

**Figure 25A-1, 25A-2, 25B, 25C:** Semi-log plots of reaction showing biphasic kinetic profiles. One of the conventional ways to plot first-order kinetic data is a semi-log plot. The reaction timecourses for self-splicing show biphasic behavior for which each phase is linear. The inflection point between the two phases of reaction occurs at approximately 50% (-0.3 on the y-axis), regardless of what type of monovalent in is added. All of the reactions shown were conducted in 100 mM $MgCl_2$ and 500 mM of the monovalent slats indicated. The inset to the pot of $(NH_4)_2SO_4$ shows a considerable burst of activity during the first 13% of reaction (-0.06 on the y-axis). Note that the data for the next phase of splicing (after the burst) is completely linear when shown in detail.

**Figure 26A, 26B, 26C:** Kinetic analysis of self-splicing in the presence or monovalent cations: The experimental data for decrease in precursor (□), and the evolution of linear (●) and lariat (▲) intron products are plotted and fit using equations 1-10 of the text. The kinetic parameters cited below and in Table 5 were determined from these fits. (A) Kinetics of splicing in 500 mM Kcl and 100 mM $MgCl_2$: From the fit of the data, the $k^A_T$ and $k^B_T$ values for the reaction of precursor population A and B were 0.028±0.0031 $min^{-1}$ and 0.0042±0.00026 $min^{-1}$, respectively. The fractional coefficients for A($f^A_{hy}$ and $^A f_{br}$) were 0.32±0.040 and 0.093±0.022, respectively. The coefficients for B($f^B_{hy}$ and $f^B_{br}$) were 0.48±0.068 and 0.16±0.073 respectively. These were used to compute the individual rate constants for branching and hydrolysis that are reported in Table 5. Values of $k^A_T$ determined from linear and lariat intron evolution were 0.019 $min^{-1}$ and 0.037 $min^{-1}$, respectively. The $k^B_T$ determined from lariat evolution was 0.0030 $min^{-1}$. Based on other estimations, the $k^B_T$ value for linear evolution was fixed at 0.0035 $min^{-1}$. (B) Kinetics of splicing in 500 mM $NH_4Cl$ and 100 mM $MgCl_2$: From the fit of the precursor data, the $k^A_T$ and $k^B_T$ values for reaction of precursor were 0.038±0.0044 $min^{-1}$ and 0.0039±0.00027 $min^{-1}$, respectively. From the fit of product evolution data, the fractional coefficients for A($f^A_{hy}$ and $f^A_{br}$) were 0.38±0.044 and zero, respectively. The coefficients for B($f^B_{hy}$ and $f^B_{br}$) were 0.10±0.018 and 0.51, respectively. The latter value was fixed based on the linear population at long timepoints. Under this reaction condition, early timepoints for linear evolution could not be accurately differentiated from background counts so they were not included. These values were used to compute the individual rates of branching and hydrolysis shown in Table 5. The value of $k^A_T$ determined from lariat formation was 0.037 $min^{-1}$. Values of $^B k_T$ determined from branching and hydrolysis were 0.0057 $min^{-1}$ and 0.0032 $min^{-1}$, respectively. (C) Kinetics of splicing in 500 mM $(NH_4)_2SO_4$ and 100 mM $MgCl_2$: From the fit of precursor

reaction using a fixed 13% burst population, the rates of the burst, $k^A_T$ and $k^B_T$ were 0.22±0.037 $min^{-1}$, 0.047±0.0033 $min^{-1}$ and 0.0028±0.00014 $min^{-1}$, respectively. The fractional coefficients for A($f^A_{hy}$ and $f^A_{br}$) were 0.39±0.0044 and zero, respectively. The coefficients for B($f^B_{hy}$ and $f^B_{br}$) were 0.22±0.050 and 0.39, respectively. The latter value was fixed based on the linear population observed at long timepoints. These values were used to compute the individual rates of branching and hydrolysis shown above and in Table 5.

Figure 27: Comparison of ai5g self-splicing products and reactions: In order to qualitatively illustrate the relative amounts of reaction products and to show the differential amounts of nonspecific RNA degradation under all the conditions described here and in the literature, short timecourses were performed and subjected to electrophoresis on a 4% denaturing high-resolution polyacrylamide gel. The wavy mobility of RNA in lanes 14-18 is due to stretching of the low-percentage gel during work-up and is not a characteristic of the indicated reactions. As in Figure 23A, the bands are assigned as follows: A, lariat intron; B, precursor; C, linear intron; D, spliced exons; E, free 3'-exon; F, free 5'-exon. Timepoints were chosen in order to show similar amounts of reaction under different conditions. Lane 1 is the precursor RNA at t=0. For lanes 2-20, the reaction time are 5, 20, 65, 900, 1.5, 20, 45, 6, 14, 50, 30, 70, 160, 35, 75, 170, 30, 70 and 160 minutes, respectively, under the reaction conditions shown. Lane M contains RNA markers of the molecular weights shown. For reactions shown in lanes 2-11, the reaction conditions included 100 mM $MgCl_2$, together with 500 mM of the indicated salt. Lanes 12-14 are self-explanatory. Lanes 15-17 include 10 mM $MgCl_2$, together with 2 mM spermidine. Lanes 18-20 include 10 mM $MgCl_2$, together with 5 nM protamine.

DETAILED DESCRIPTION OF THE INVENTION

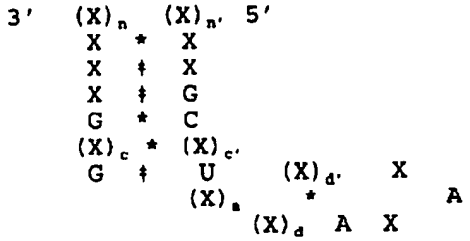
This invention is directed to a composition for catalyzed oligonucleotide cleavage comprising a synthetic non-naturally occurring oligonucleotide compound. The compound comprises nucleotides whose sequence defines a conserved group II intron catalytic region and nucleotides whose sequence is capable of hybridizing with a predetermined oligonucleotide target sequence to be cleaved, such target sequence not being present within the compound. The composition also includes an appropriate oligonucleotide co-factor which may be provided in cis or trans. Preferably, the conserved group II intron catalytic region is a group II intron domain I catalytic region. In one embodiment the conserved group II intron domain I catalytic region may further comprise a conserved portion of a group II intron domain II, a group II intron domain III, a group II intron domain IV, a group II intron domain V, or a group II intron domain VI.

In another embodiment the conserved group II intron domain I catalytic region may further comprise a portion of a group II intron sufficient to maintain catalytic activity. The conserved portions of the domain I region may comprise structural features or conserved secondary structural features sufficient for such activity. As used herein, "conserved structural features" include the EBS1 domain, EBS2 domain, $\alpha$1 domain, $\alpha'$ domain, $\epsilon$ domain, $\epsilon'$ domain, and the bulge between the D domain and the D' domain. The sequenes of the EBS1 and EBS2 domains may be modified to hybridize with the desired target sequence, so long as sufficient secondary structure is preserved.

In the composition above, the nucleotides whose sequence is capable of hybridizing with a predetermined oligonucleotide target sequence to be cleaved may be two hybridizing regions, each region having 2 to 12 nucleotides capable of hybridizing with the oligonucleotide target to be cleaved.

In one embodiment, each of the regions has 6 or 7 nucleotides. Preferably, the synthetic non-naturally occurring oligonucleotide compound comprises between 80 and 1000 nucleotides. In one embodiment the compound comprises 300-600 nucleotides.

In one embodiment of the invention, the appropriate co-factor is an oligonucleotide compound from a conserved portion of a group II intron domain V having the formula:




wherein each X represents a ribonucleotide which may be the same or different; wherein each of $(X)_n$, $(X)_{n'}$, $(X)_c$, $(X)_{c'}$, $(X)_{d'}$, $(X)_a$ and $(X)_d$ represents an oligonucleotide; wherein n, n', c, c', d, d', and a each represents an integer which defines the number of nucleotides in the oligonucleotide with the provisos that n and n' are greater than or equal to 1; a represents an integer which is greater than or equal to 1; d and d' represent an integer which is greater than or equal to 5; c and c' represents an integer which is greater than or equal to 4; wherein each * represents base pairing between the nucleotide located on either side thereof; wherein each † may or may not represent base pairing between the nucleotide located on either side thereof; and wherein each solid line represents a chemical linkage providing covalent bonds between the nucleotide located on either side thereof. In one embodiment, 5' $(X)_{n'}$XX may be 5' $(X)_{n'}$XA.

The compositions above may further comprise a divalent cation such as $Mg^{2+}$ or $Mn^{2+}$. Preferably, the predetermined oligonucleotide target sequence to be cleaved and/or ligated is mRNA e.g. a mammalian mRNA, a yeast mRNA, a bacterial

mRNA or a viral mRNA. However, the predetermined oligonucleotide target sequence to be cleaved may be DNA. The mRNA may encode a growth factor such as an angiogenic factor, a basic fibroblast growth factor, a colony-stimulating factor 1, cystic fibrosis transmembrane conductance regulator, an epidermal growth factor, an erythropoietin, a fibroblast growth factor, a G-protein, a granulocyte-macrophage colony stimulating factor, a growth hormone, IL-1, IL-1R, IL-2, IL-2R, IL-4, IL-6, an insulin-like growth factor, an insulin-like growth factor 1, an interferon, an interleukin, a keratinocyte growth factor, luteinizing hormone receptor, MDR-1, a nerve growth factor, a platelet derived growth factor, a scatter factor, a transforming growth factor $\alpha$, a transforming growth factor $\beta$, a transforming growth factor, or a tumor necrosis factor.

The mRNA may encode an oncogene or a tumor suppressor gene such as *bcl*-2, *bcr-abl*, *bek*, BPV, c-*abl*, c-*fes*, c-*fms*, c-*fos*, c-H-*ras*, c-*kit*, c-*myb*, c-*myc*, c-*mos*, c-*sea*, *cerbB*, DCC, erbA, *erbB*-2, *ets*, *fig*, FSFV gp55, Ha-*ras*, HIV *tat*, HTLV-1 *tat*, JCV early, *jun*, L-*myc*, *lck*, LPV early, *met*, N-*myc*, NF-1, N-*ras*, *neu*, p53, Py mTag, *pim*-1, *ras*, RB, *rel*, retinoblastoma-1, SV-40 Tag, TGF-$\alpha$, TGF-$\beta$, *trk*, *trkB*, *v-abl*, v-H-*ras*, *v-jun*, or WT-1. The mRNA may be associated with a chromosomal translocation or have one or more point mutations. In addition, the mRNA may be an mRNA whose overproduction is associated with a disease or condition.

The mRNA may be a viral mRNA associated with a cytomegalovirus, an Epstein-Barr virus, a hepatitis B virus, a hepatitis C virus, a herpes simplex virus, a herpesvirus, a HIV-1 virus, an immunodeficiency virus, an influenza virus, a papillomavirus, a picornavirus, a polio virus or a T-cell leukemia virus. Alternatively, the mRNA may be a plant mRNA associated with alfalfa, apples, asparagus, bananas, broccoli, carrots, celery, chicory, coffee, cabbage, mustard, corn, cottonseed, squash, cucumber,

cantaloupe, grapes, lettuce, palm, potato, rapeseed, raspberry, soybean, sunflower, strawberry, tomato, or wheat.

The invention is also directed to an oligonucleotide transfer vector containing a nucleotide sequence which on transcription gives rise to the synthetic non-naturally occurring compounds above. The transfer vector may be a bacterial plasmid, a bacteriophage DNA, a cosmid, or an eukaryotic viral DNA.

The invention is also directed to host cells transformed by the transfer vector above. The host cells may be prokaryotic host cells or an eukaryotic host cells such as E. coli, a monkey COS host cell, a Chinese hamster ovary host cell, a mammalian host cell or a plant host cell.

The invention is also directed to a method for cleaving an oligonucleotide comprising the steps of: a) providing a synthetic non-naturally occurring oligonucleotide compound or compounds as defined above and b) contacting the synthetic non-naturally occurring oligonucleotide compound(s) with a target oligonucleotide to cause cleavage of the target oligonucleotide.

A transgenic, non-human vertebrate animal, having one or more cells bearing a DNA sequence encoding the composition above. The transgenic animal may be a mammal such as a cow, goat, sheep, pig, horse, dog, cat, or rodent. In addition, the animal may be a fish, a chicken, or a turkey.

The invention is also directed to a method of treatment which comprises: a) recovering a subject's appropriate cells; b) transforming the cells with a DNA sequence encoding the composition of above; and c) introducing the resulting transformed cells into the subject so as to treat the subject.

Synthetic preparations of mRNA are well known (see Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd Ed.). Mixed DNA-RNA oligomers with modified base pairs for the ribozyme or minizyme can be prepared by commercially available DNA synthesizers such as those produced by Applied Biosystems. For derivatives see (Uhlmann, E. and Peyman, A., 1990, Chemical Reviews 90:543-584.), for H-phosphonate monomers (Agrawal et al U.S. Patent No. 5,149,798).

The compounds of this invention may be covalently or non-covalently associated with affinity agents such as proteins, antibodies, steroids, hormones, lipids, specific nucleic acid sequences, intercalating molecules (such as acridine derivatives, for example 9-amino acridine) or the like to modify binding affinity for a substrate nucleotide sequence or increase affinity for target cells, or localization in cellular compartments or the like. For example, the compounds of the present invention may be associated with RNA binding peptides or proteins which may assist in bringing the endonuclease into juxtaposition with a target nucleic acid such that hybridization and cleavage of the target sequence may take place. Nucleotide sequences may be incorporated into the 5' and 3' ends of the group II intron increase affinity for substrates. Such additional nucleotide sequences may form triple helices with target sequences (Strobel, S.A., et al., 1991, Nature 350:172-174.) which may enable interaction with intramolecularly folded substrate. Alternatively, modified bases within the additional nucleotide sequences may be used that will associate with either single stranded or duplex DNA generating base pair, triplet, or quadruplet, interactions with nucleotides in the substrate.

The compounds of the claimed invention may be further stabilized using methods in the literature for example the use transcription terminators on the 3' end such as the T7 terminator, $\rho$-independent terminator, cry element (Gelfand

et al. U.S. Patent No. 4,666,848) or the TrpE terminator. Furthermore, sequences such as the poly(A) addition signal AATAAA may be added and strategies involving changing the length of the 3' non-coding region (see Gillies, U.S. Patent No. 5,149,635). These techniques can be used to stabilize RNA in the compound.

The invention also embodies methods of production of the nucleic acid based compounds described above comprising the steps of: (a) ligating into a transfer vector comprised of DNA, RNA or a combination thereof a nucleotide sequence corresponding to said compound; (b) transcribing the nucleotide sequence of step (a) with RNA polymerase; and (c) recovering the compound. The invention also includes transfer vectors, bacterial or phage, comprised of RNA or DNA or a combination thereof containing a nucleotide sequence which on transcription gives rise to the compounds described above.

The invention described herein also provides a method of cleavage of a specific RNA target sequence which comprises reacting the compound with the target sequence so as to thereby cleave the specific target sequence. Such target sequences may be indigenous to mammals or plants. In one embodiment, the target sequence is in a viral gene. The invention also provides a method for the treatment of viral diseases in plants and animals.

Further, many methods have been developed for introducing cloned eukaryotic DNAs into cultured mammalian cells (Sambrook et al., 1989): calcium phosphate- or DEAE-dextran-mediated transfection; polybrene; protoplast fusion; electroporation; and direct microinjection into nuclei. The present invention, however, extends to other means of transfer such as genetic bullets (e.g. DNA-coated tungsten particles, high-velocity micro projectile bombardment) and electroporation amongst others (Maliga, P. (1993) Towards

plastid transformation in flowering plants. Tibtech 11:101-106; Bryant, J. (1992) Transgenic wheat plants: the end of the beginning. Tibtech 10:342-343.; or Shimamoto, K., Terada, R., Izawa, T., and Fujimoto, H. (1989) Fertile transgenic rice plants regenerated from transformed protoplasts. Nature 338:274-276.).

Further, the compound described herein may be used in plants to cleave undesirable mRNA. The appropriate cleavage would lead to phenotypic changes. Phenotypic changes in plant cells or plants may include drought resistance, salinity resistance, resistance to fungal, viral or bacterial infection; modifications of growth characteristics; sterility; fruit production; flowering; senescence; altering oil seed metabolic pathways to increase production; and the like (see Shewmaker et al. U.S. Patent No. 5,107,065). It is evident that one or more RNA involved in determining phenotype are identified, such RNAs may be inactivated by cleavage utilizing the endonuclease of this invention and thus the phenotype of the plant or plant cell altered. Diseases or infections which may be treated in plants with endonucleases of this invention include fungal infection, bacterial infections (such as Crown-Gall disease) and disease associated with plant viral infection.

Phenotypic modifications within animals (including in some applications man) which may be effected by cleaving and thus inactivating target RNAs associated with phenotype would include growth characteristics of animals, fertility, skin/cosmetic modifications, reproductive characteristics, disease resistance and the like. A myriad of applications arise for phenotypic modifications in animals, and plants as previously mentioned. One or more RNAs associated with a given endonucleases may be targeted against such RNAs for their inactivation with consequential phenotypic modification.

Prokaryotic and eukaryotic cell cultures may be phenotypically modified by treatment with endonucleases of this invention. For example, bacterial cultures or yeast cultures involved in production of food components (such as cheese, bread and dairy products) and alcoholic beverage production may be treated so as to modify enzyme content, flavor production, cell growth rate, culture conditions and the like. Eukaryotic and prokaryotic cells in culture may, for example be protected from infection or disease associated with mycoplasma infection, phage infection, fungal infection and the like.

The compounds of this invention may also be used to treat diseases or infection in humans, animals, plants, or prokaryotic or eukaryotic cells. The ability to treat disease or infection is based on the fact that the compounds of this invention are capable of cleaving any RNA which contains a suitable cleavage site. Most RNAs will contain one or more suitable cleavages sites.

The period of treatment would depend on the particular disease being treated and could be readily determined by a physician. Generally treatment would continue until the disease being treated was ameliorated.

Examples of human and animal targets may be bacterial and prokaryotic infection, viral infection and neoplastic conditions associated with the production of aberrant RNAs such as occurs in chronic myeloid leukemia, viruses, growth factors as described above.

The virus may be a plant virus such as a tobamovirus, a tobravirus, a hordeivirus, a potexvirus, a carlavirus, a potyvirus, a closterovirus, a tymovirus, a tombusvirus, a sobemovirus, or a luteovirus. The plant virus may be a potato yellow dwarf virus, a cucumber mosaic virus, a tomato spotted wilt virus, a tomato mosaic virus, a potato virus X

(PVX), a potato virus Y (PVY), a carnation latent virus, a tomato rattle virus, a pea early browning virus, a barley stripe mosaic virus, a turnip yellow mosaic virus, a barley yellow dwarf virus, a beet yellows virus, a potato leaf roll virus, a tomato bushy stunt virus, a southern bean mosaic virus, a maize chlorotic virus, beet necrotic yellow vein virus, or a tobacco necrosis virus.

The transgenic organisms described herein may be characterized in that they contain in their genome a sequence which gives rise, on transcription, to the compounds mentioned above. The transgenic organism may be a plant or a non-human mammal. Examples of transgenic mammals include Leder et al. U.S. Patent Nos. 4,736,866 and 5,175,383; Krimpenfort et al. U.S. Patent No. 5,175,384; Wagner et al. U.S. Patent No. 5,175,385; and U.S. Patent Nos. 5,183,949 and 5,347,075. The transgenic plant, including fruits, and seeds thereof, may be from alfalfa, apple, bean, canola (oilseed rape), cantaloupe, corn, cotton, courgette, cucumber, melon, papaya, pepper, potato, rice, soybean, squash, strawberry, sunflower, sweet pepper, tobacco, tomato, or walnut. Also included are the plant cells transformed by the above-mentioned transfer vector, as well as a prokaryotic or eukaryotic cell, plant or animal, comprising a nucleotide sequence which is, or on transcription gives rise to, the compounds described herein.

An effective amount of a compound of the present invention would generally comprise from about 1 nM to about 1 mM concentration in a dosage form, such as a cream for topical application, a sterile injectable composition, or other composition for parenteral administration. In respect of topical formulations, it is generally preferred that between about 50 $\mu$M to about 500 $\mu$M endonuclease be employed. Compounds comprising nucleotide derivatives, which derivatives may involve chemically modified groups, such as phosphorothioate or methyl phosphonate derivatives may be

active in nanomolar concentrations. Such concentrations may also be employed to avoid toxicity.

Therapeutic strategies involving treatment of disease employing compounds of this invention are generally the same as those involved with antisense approaches, such as described in the antisense literature (Chrisley, L.A. (1991) Antisense Research and Development, 1:65-113; Stein, C.A. et al. Science (1993) 261:1004-1012). Particularly, concentrations of compounds utilized, methods and modes of administration, and formulations involved may be the same as those employed for antisense applications.

An "effective amount" as used herein refers to that amount which provides a desired effect in a mammal having a given condition and administration regimen. Compositions comprising effective amounts together with suitable diluents, preservatives, solubilizers, emulsifiers, adjuvants and/or carriers useful for therapy. Such compositions are liquids or lyophilized or otherwise dried formulations and include diluents of various buffer content (e.g., Tris-HCL, acetate phosphate), pH and ionic strength, additives such as albumin or gelatin to prevent absorption to surfaces, detergents (e.g., Tween 20, Tween 80, Pluronic F68, bile acid salts), solubilizing agents (e.g., Thimerosal, benzyl alcohol), bulking substances or tonicity modifiers (e.g., lactose, mannitol), covalent attachment of polymers such as polyethylene glycol to the oligonucleotide, complexation with metal ions, or incorporation of the material into or onto particulate preparations of polymeric compounds such as polylactic acid, polyglycolic acid, polyvinyl pyrrolidone, etc. or into liposomes, microemulsions, micelles, unilamellar or multilamellar vesicles, erythrocyte ghosts, or spheroplasts. Such compositions will influence the physical state, solubility, stability, rate of in vivo release, and rate of in vivo clearance of the oligonucleotide. Other ingredients

optionally may be added such as antioxidants, e.g., ascorbic acid; low molecular weight (less than about ten residues) polypeptides, i.e., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; amino acids; such as glycine, glutamine acid, aspartic acid, or arginine; chelating agents such as EDTA; and sugar alcohols such as mannitol or sorbitol. Possible sustained release compositions include formulation of lipophilic depots (e.g., fatty acids, waxes, oils). Also comprehended by the invention are particulate compositions coated with polymers (e.g., polyoxamers or polyoxamines) and oligonucleotides coupled to antibodies directed against tissue-specific receptors, ligands or antigens or coupled to ligands of tissue-specific receptors. Further, specific nucleotide sequences may be added to target the oligonucleotides of this invention to the nucleus, cytoplasm or to specific types of cells. Other embodiments of the compositions of the invention incorporate particulate forms protective coatings, protease inhibitors or permeation enhancers for various routes of administration, including parenteral, pulmonary, nasal and oral.

Suitable topical formulations include gels, creams, solutions, emulsions, carbohydrate polymers, biodegradable matrices thereof; vapors, mists, aerosols, or other inhalants. The oligonucleotides may be encapsulated in a wafer, wax, film or solid carrier, including chewing gums. Permeation enhancers to aid in transport to movement across the epithelial layer are also known in the art and include, but are not limited to, dimethyl sulfoxide and glycols.

For in-vitro use, the compounds of this invention are generally reacted with a target RNA which contains one or more suitable cleavage sites. Optionally, the target RNA may be purified or substantially purified. The nucleotide sequences the compounds of this invention are selected so as to specifically hybridize or form a double-stranded duplex

with a target RNA whereafter cleavage takes place. Accordingly, target RNA may be specifically cleaved in-vitro in the presence of other RNAs which themselves would not be cleaved.

The compounds may be utilized in a manner similar to restriction endonucleases, that is for the specific cleavage of RNA to facilitate RNA manipulation. All that is required for such manipulations is that the target RNA to be cleaved contains a uracil base and thus a suitable cleavage site.

The compounds of this invention may be utilized in diagnostic procedures, such as the mapping or finger-printing of RNA. Specifically, the compounds of this invention would enable mapping of RNA and may be used to detect mutations in RNA sequence. Such procedures may be used in research and may also have forensic and other diagnostic applications.

RNA cleavage products in-vitro may be readily detected, for example, by visualization on acrylamide or agarose gels where the amounts of RNA cleaved are sufficiently large for direct visualization after separation and reaction with nucleotide visualization agents, such as ethidium bromide. Alternatively, where the target RNA cleaved is present in small amounts, such as in a sample containing many RNAs, cleavage products may, for example, be detected by using radiolabelled probes for sequence complementary to the target sequence, or amplification techniques such as PCR (Sambrook et al., 1989).

A target RNA for cleavage in-vitro may be derived from any source, and may be of animal, viral, bacterial, plant, synthetic, or other origin. As RNA is common to all known living organisms, this invention may be utilized to cleave any RNA species having a suitable cleavage site as mentioned previously.

This invention is illustrated in the Experimental Detail sections which follow. These sections are set forth to aid in an understanding of the invention but are not intended to, and should not be construed to, limit in any way the invention as set forth in the claims which follow thereafter.

## EXPERIMENTAL DETAILS

### Example 1:

A self-splicing group II intron was transformed into a three-part ribozyme that site-specifically cleaves small oligonucleotide substrates with multiple turnover. The multi-component structure of the ribozyme facilitated quantitation of individual rate constants, interactions between reaction components (5'-exon, Domain 1 and Domain 5), and reaction stereospecificity. Domain 1 and Domain 5 may or may not be located on the same molecule. The efficient reaction catalyzed by this ribozyme appears to be limited by the chemical rate of hydrolysis. An Rp phosphorothioate substrate may be the preferred diastereomer of the ribozyme.

Self-splicing group II introns are essential for the processing of many organella genes in plants and yeast (Umesono, 1989; Kück, 1990). Group II introns are characterized by a conserved secondary structure that can be organized into six domains. The splicing reaction is initiated by attack of a 2'-hydroxyl group projecting from the bulged adenosine of Domain 6(D6), which attacks the 5'-splice site and generates a "lariat" intermediate (Peebles, 1986; van der Veen, 1986).

In group II introns, both steps of splicing can proceed if the initial nucleophile is a water molecule rather than the 2'-OH of a bulged adenosine in D6 (Jarrell, 1988(a); Jarrell, 1988(b); van der Veen, 1987). In addition,

hydrolysis promotes the apparent first step of group II intron trans-splicing by 5'-exon-G oligonucleotides (Jacquier, 1986; Altura, 1989). Domain 5 (D5) is the most phylogenetically conserved region of the group II intron and it has been observed to act in-trans, promoting specific hydrolytic cleavage tat the 5'-splice site (Jarrell, 1988). It is clear that D5 and Domain 1 (D1) are absolutely required for the first step of splicing (Koch, 1992). Recent kinetic analysis have examined the hydrolysis of 5'-exon from domains 1-3 when D5 is provided in-trans (Franzen, 1993; Pyle, 1994). These studies showed that D5 associates with high affinity (~300 nM) to the other active site components. The strength of D5 binding is especially surprising considering that this small RNA (~40 nts) is largely duplex in form and, having no apparent phylogenetic covariation with the rest of the intron, seems to bind without the aid of base-pairing interactions. This D5-catalyzed reaction obeys a straightforward Michaelis-Menten mechanism in which the chemical step of hydrolysis appears to be rate-limiting (Pyle, 1994).

Minimal Group II introns that retain substrate binding and catalytic activity

Further deletion studies have demonstrated that group II introns will retain substrate binding capability and catalytic activity without certain subdomains. Please refer to the secondary structure of the yeast mitochondrial group II intron depicted in Figure 2A,2B. These experiments have shown that the group II intron will remain fully active for splicing with the removal of the D2b domain, the C2 domain, and the lower stem of the C1 domain. Additionally, domain II and domain IV can be completely deleted and splicing activity of the molecule will remain intact. We have demonstrated that the combination of domain III and domain I results in a very efficient ribozyme. Domain III appears to speed up the kinetics of the chemical steps involved in splicing.

In order to independently evaluate the role of D1, D5 and 5'-exon sequences on active-site folding and catalytic activity, we constructed a modular group II intron ribozyme consisting of a 425 nt Domain 1 RNA (D1) that binds small oligonucleotide substrates (s, 24-27 nts) in trans. Efficient cleavage at a single site is promoted by addition of a reaction cofactor (D5, 58 nts) that is also provided in trans (Fig. 4a). This ribozyme is analogous to the Tetrahymena ribozyme except that the cofactor is a short RNA which is unchanged in the reaction instead of a nucleophilic guanosine that is consumed during reaction (Zaug, 1986 (a); Zaug, 1986(b); Pyle, 1994). With the reaction substrate and cofactors cleanly separated, it is now possible to hold any one component rate limiting in order to determine its precise role in binding or catalysis. The modular structure permits facile incorporation of mutations and single-atom changes for probing mechanism (Cech, 1992). We have used this true ribozyme to begin construction of a more complete kinetic framework for group II intron catalysis and to determine the stereospecificity of a reaction analogous to the first step of group II intron self-splicing.

**Kinetics of Substrate Cleavage by the Ribozyme:** The reaction substrate (S) is a small oligonucleotide with a sequence analogous to the 5'-exon/intron boundary in the ai5g group II intron (Fig. 4a). In the presence of D1 and D5, this oligonucleotide is cleaved specifically at the CpG sequence that corresponds to the normal 5'-splice site (Fig. 4b). If either D1 or D5 is absent from the reaction, no cleavage of the substrate occurs. Previous studies showed that a 52-nt 5'-exon terminating in the CpG cleavage site was the minimal substrate necessary for trans-splicing or trans-cleavage by the ai5g group II intron (Jacquier, 1986; Jacquier, 1987). In our study, only 17 nts of exon, encompassing both IBS1 and IBS2 (Fig. 4) were required for efficient cleavage. This substrate also contains 7-8

nucleotides of intronic sequence, including nucleotides (+3 and +4) involved in the ∈-∈′ interaction believed to stabilize 5′-splice-site association (Jacquier, 1990).

$K_{ms}$ $k_{max}$ and $k_{max}/K_m$ values for S were obtained using single-turnover kinetics where [S] << $K_{ms}$ and [D5] was saturating relative to [D1] (Table 1, line 1) according to standard kinetic procedures (Fig. 5, Table 1, line 1) (Pyle, 1994; Herschlag, 1990; Fedor, 1992). $K_m^S$ determined from this plot is 6.3 Nm, implicating strong association between S and D1 (see IBS/EBS pairings, Fig. 4a). This is the first quantitative analysis of the strength of binding between the 5′-exon and D1, and the $K_m^S$ approximates a dissociation constant for the IBS/EBS pairings plus any other interactions mediating intron/exon association. At very high concentrations of D1 (the plateau of the plot in Fig. 5A) $k_{obs}$ reflects the conversion of E·S → E·P, where E represents D1 that is fully saturated with D5 cofactor and P represents reaction products. Under these saturating, single-turnover conditions, the apparent rate (defined here as $k_{max}$) represents the rate of the chemical step and any conformational changes that take place within the E·S complex. The $k_{max}$ value at saturating [D1] was 0.021 $min^{-1}$.

Table 1 - Kinetic Parameters for Substrate Cleavage by the Group II Introns

| Substrate | rate-lim component | $k_{max}$ $(min^{-1})^{\S}$ | $K_m$ (nM) $^{\S}$ | $k_{max}/k_m$ $(M^{-1}min^{-1})^{\Psi}$ | $k_{max}$ $(min^{-1})^{\S}$ | thio |
|---|---|---|---|---|---|---|
| $S_{(synthetic)}$ | D1 | 0.21± 0.001 | 6.3± 0.88 | (3.3±0.49) X $10^6$ | 0.012 | 1.8 |
| $S_{(synthetic)}$ | D1 | 0.032± 0.0048 | - | - | 0.0094 | 3.4 |
| $S_{(synthetic)}$ | D5 | 0.019± 0.0010 | 870± 110 | (2.2±0.30) X $10^4$ | - | - |

Table 1 legend:

' Described in legend to Fig. 5. $^{\S}$ $k_{max}$ represents the maximum rate at saturation and represents the horizontal asymptote derived from the binding curves shown in Fig. 5. Theoretical $k_{max}$ values at ωD1 and ωD5 were calculated from the data shown in Fig. 5 and found to be in agreement with a value of 0.026 $min^{-1}$ determined experimentally at 100 nM D1 and 9 μM D5. $K_m$ represents the apparent binding constant of the limiting component. These values were determined from the fit of the curves shown in Fig. 5 to an equation describing 1:1 bimolecular association[12]. The standard error for the fit is shown. The appropriateness of the standard error calculation was confirmed using a jackknife approximation of standard error[12]. $^{\Psi}$ $k_{max}/K_m$ (analogous to $k_{cat}/K_m$ values determined from the slopes of $k_{obs}$ vs. [D1] or [D5] at very low concentrations. Reported error was propagated from the $k_{max}$ and $K_m$ standard errors. ' The reported rates represent $k_{obs}$ values for phosphorothioate substrates at saturation, where [D1] = 100 nM and [D5] = 3 μM (analogous to the plateau of the curve in Fig. 5A). Each value is the average of two trials. Like all reported $k_{obs}$, variance is 18% (see legend to Fig. 5A). The empirically

determined endpoint (after 7 reaction half-times) used in calculation of $k_{obs}$ was 82% for the transcribed substrate and 48% for the synthetic substrate. .. The ratio of $k_{max}$ to $k_{obs}$ for phosphorothioate substrates at saturation. '' This value is the $k_{obs}$ at D1 and D5 saturation for cleavage of a transcribed all-phosphate substrate.

For conditions under which [D5] was rate-limiting, values for $K_m^{D5}$, $k_{max}^{D5}$ and $k_{max}/K_m^{D5}$ were determined by monitoring the cleavage rate of S in the presence of saturating D1 through a range of D5 concentrations (Fig. 5B, Table 1, line 3). In this case, each molecule of S was bound to D1 and overall rate was limited by the concentration of D5. The $k_m^{D5}$ is 870 nM, $k_{max}/K_m^{D5}$ is 2.2 x $10^{-1}$ M $min^{-1}$ and $k_{max}$ is 0.019 $min^{-1}$. Values of $k_{max}$ from both D1 and D5-limited experiments are in close agreement (Table 1). The consistency of the data obtained under limiting D1 and D5 conditions suggests that under either condition, reaction proceeds through the same ternary complex for which the individual components have different affinities.

Specific, efficient cleavage of small substrates by the group II ribozyme has now been observed in a kinetically characterized system. The results are consistent with kinetic behavior of larger, more complex constructs of the group II ribozyme has now been observed in a kinetically characterized system. The results are consistent with kinetic behavior of larger, more complex constructs of the group II intron (Jarrell, 1988; Pyle, 1994). The ribozyme construct can now be used to carry out detailed enzymological studies that examine reaction mechanism at an atomic level. Of particular interest is the stereospecificity of group II catalysis of the first and second steps of splicing (Lamond, 1993; Moore, 1993(a); Moore, 1993(b)). Group II introns are invoked as an autocatalytic model for the spliceosome and there are indications that group II introns and pre-mRNA splicing

share a common evolutionary ancestry (Cavalier-Smith, 1991).

**Stereospecificity of Cleavage by a group II Ribozyme:** Synthetic incorporation of a chiral phosphorothioate linkage results in an equal mix of Rp and Sp diastereomers. A synthetic substrate for the ribozyme was machine-synthesized, containing a single phosphorothioate linkage at the cleavage site. The ribozyme cleaves this substrate to an endpoint of ~50%, suggesting that it has a preference for one of the two diastereomers (Fig. 6). However, the fraction uncleaved by the ribozyme could represent RNA that is unreactive for other reasons, such as the presence of protecting groups left over from chemical synthesis. To determine the stereochemical identity of the fraction uncleavable by the ribozyme, the uncleaved fraction was gel-purified and examined by digestion with snake venom phosphodiesterase (SVPD). SVPD is a 3'→ 5' exonuclease that readily cleaves phosphate and Rp phosphorothioate linkages. This enzyme is markedly inhibited by Sp phosphorothioates, cleaving them 2000X more slowly than their corresponding Rp isomers (Burgers, 1979). When a 5'-end labelled RNA contains an internal Sp phosphorothioate linkage, SVPD will readily digest the 3'-end of the strand and suddenly pause at the site immediately adjacent to the Sp linkage (+1G). This apparent stall will accumulate and then diminish at long times as the SVPD slowly cleaves through it and then rapidly digests the remaining strand to mononucleotides. Three 5'-end labelled RNAs were subjected to cleavage by SVPD (Fig. 7). An all-phosphate substrate was readily digested, with no apparent stalling at position 18, adjacent to the ribozyme cleavage site (Fig. 7, phosphate panel). Digestion of the racemic substrate (never treated with ribozyme) resulted in prominent exonucleolytic stalling adjacent to the single phosphorothioate incorporated at the ribozyme cleavage site. After incubating the racemic substrate with ribozyme (for > 7 reaction half-times), the ribozyme-uncleavable fraction was digested with SVPD to

yield a predominant stall product that was much greater in intensity than observed for the racemic mix (Fig. 7, Sp panel). For each lane, the number of counts in the stall band were normalized for the total number of counts in the entire lane (from 1 → 24 nts). Using this value, we calculated the integrated population of the stall product over the duration of the timecourses shown. The integrated intensity of the stall product in the "Sp" panel (Fig. 7) was 1.0 relative to the integrated intensity of 0.59 in the racemic "Rp + Sp" panel. An integrated intensity 0.05 for the "phosphate" panel may be taken as a measure of background. Together with the fact that the racemic substrate is cleaved by the ribozyme to ~50%, these results suggest that the group II ribozyme may preferentially cleave the Rp diastereomer.

Taking into account the 50% endpoint for cleavage of the mixed substrate, rate constants for cleavage of the phosphorothioate substrate were readily obtained (Table 1). A two-fold diminution in rate (the "thio-effect") was observed for cleavage of a phosphorothioate linkage. The magnitude of this rate-decrease, upon substitution of phosphate with phosphorothioate, has been cited as evidence for a reaction rate limited by chemistry rather than conformational changes (Herschlag, 1991). This would be consistent with results from previous studies of 5'-splice-site hydrolysis by a two-piece ribozyme. Correspondence between multiple-turnover and single-turnover kinetic parameters, together with the observations that $K_m^{DS} \approx K_d^{DS}$ and that rate was log-linear with pH led to the hypothesis that the reaction was limited by the chemical rate of hydrolysis (Pyle, 1994).

In order to directly test the hypothesis that the ribozyme cleaves an Rp diastereomer, a new substrate was prepared by in-vitro transcription. T7 RNA polymerase incorporates only Rp phosphorothioates into RNA polymers (Griffiths, 1987).

Using α-thio-GTP, we transcribed an RNA substrate containing an Rp phosphorothioate at every G residue, including the 5'-splice site (5'SS). Just as the ribozyme specifically cleaves a phosphodiester linkage at the 5'SS sequence (Fig. 8, lane A8), the ribozyme specifically cleaves a substrate containing an Rp phosphorothioate linkage at the cleavage site (Fig. 8, lane B8). To confirm that the ribozyme cleaves both phosphodiester and phosphorothioate RNA substrates uniquely at the 5'SS sequence, the transcribed substrates and products were mapped with a battery of sequence- and stereo-specific nucleases (Fig. 8, lanes 4-7, 9). The substrate containing all-phosphate linkages was mapped next to the substrate containing phosphorothioates at G resides (Fig. 8, Sections A and B). The proper sequence at the cleavage sites was confirmed by digestion of substrates with C13, B-cereus and T1 endonucleases (Boguski, 1979; Donis-Keller, 1977). Product oligonucleotides migrate at the expected position relative to specific nuclease cuts at -1C and +1G on the substrates. The endonucleolytic digests described above, together with U2 digestion, confirmed the absolute sequence throughout both substrates (Fig. 8). The migration of the ribose product with a 19-mer P1 endonuclease digestion product suggests that the group II ribozyme leaves 3'-OH and 5'-phosphate termini. The presence of a phosphorothioate at the cleavage site was confirmed by a gap in the P1 nuclease ladder of the thio-substrate (lane 9), together with $I_2$ cleavage of the thio-substrate ($i_2$ data not shown) (Potter, 1983; Schatz, 1991). Taken together, these data show that the ribozyme is selecting the proper cleavage site on both ribose and phosphorothioate-containing substrates. Thus, the ribozyme readily cleaves the Rp diastereomer of a transcribed substrate and does not select adjacent phosphate sites for attack.

Even though the ribozyme specifically cleaves transcribed substrate at an Rp phosphorothioate, the partial digests do

not provide quantitative evidence that the substrate sequence and linkage sulfur content are homogeneous at that site. Significant de-thioation at the splice-site linkage would result in conversion to phosphate and apparently efficient cleavage by the ribozyme. We sought a quantitative demonstration that the group II ribozyme cleaves a linkage that is uncleavable by P1 endonuclease. To address these concerns, the thio-substrate was transcribed in the presence of $\alpha$-$^{32}$P-CTP. Because P1 endonuclease cannot cleave RP phosphorothioates, a total P1 digest of this substrate leads to three products in approximately equimolar ratios: *pC, *pCp$_s$G and *pCpsGp$_s$G where p$_s$ indicates a phosphorothioate linkage (Fig. 9A, lane 3). Because the intensity of the three bands are approximately equimolar, this indicates that the substrate has been faithfully transcribed by T7 RNA polymerase and that the cleavage site dinucleotide must be an Rp linkage with the same level of phosphorothioate incorporation as positions +5G and +6G (which end up in the trinucleotide). When a sample of internally $^{32}$P-labelled substrate is cleaved to completion by the ribozyme and then immediately digested with P1 under similar conditions, two predominant bands are observed: *pC and *pCp$_s$Gp$_s$G (Fig. 9A, lane 4). The middle p*Cp$_s$G band, corresponding to the ribozyme cleavage site, is considerably diminished in intensity while the *pC band is enhanced. Incubation of substrate with d1 or D5 alone, followed by P1 digestion does not result in a diminution of the middle *pCp$_s$G band (Fig. 9A, lanes 4+5). These results quantitatively demonstrate that this group II ribozyme cleaves an Rp phosphorothioate in a transcribed substrate that is homogeneous in sequence and linkage at the cleavage site.

For the transcribed phosphorothioate substrate, the cleavage kinetics were quantitated at D1 and D5 saturation, resulting in an apparent $k_{max}$ value approximately 3-fold slower than that of a phosphate at the ribozyme cleavage site (Fig. 9B).

Like the "thio-effect" observed using synthetic substrates, this value is consistent with an apparent rate-limiting chemical step (Table 1, column 7). The rate of cleavage in this homogeneously Rp case is the same (within error) as that of the 50% synthetic substrate that is cleavable by the ribozyme. This provides further evidence that the cleaved fraction in the synthetic experiments is predominantly Rp in configuration.

The new group II ribozyme described here has provided a quantitative description of inter-domain interactions, individual rate constants and reaction stereospecificity. This is the first example of a group II intron that has been transformed into a ribozyme construct that catalyzes an efficient, single reaction on small oligonucleotide substrates. The ribozyme acts as a sequence-specific RNA endonuclease and consists of two RNA subunits which are essential for catalysis. Kinetics for the reaction are unusually clean and suggest a stable, homogeneous combination of reaction components. Like the Tetrahymena ribozyme, the design of this group II construct facilitates biochemical and kinetic studies required for identification of active-site residues and tertiary interactions involved in catalysis.

The new three-part system shows that interactions between Domain 1 and the 5'-exon are very strong and that the EBS-IBS pairings plus the first several intronic nucleotides are all that is required for 5'-splice site recognition in this reaction. Although compensatory base substitutions have been performed, this is the first study to quantitate the magnitude of 5-exon binding tot he intron (~ 6 nM) (Jacquier, 1987). The $k_{cat}/Km^S$ value for this ribozyme, (3.3 x $10^6$ $M^{-1}$ $min^{-1}$) is ~20X slower than that of the Tetrahymena ribozyme, which is rate-limited by duplex association rather than chemical or conformational steps (Herschlag, 1990). A slow chemical step (0.02-0.1 $min^{-1}$) appears to limit the rate

of RNA hydrolysis by ai5g group II ribozymes and as a result, $k_{max}/Km$ is slower than the rate of duplex formation. The results using this construct confirm that the interaction of D5 with the other components is very strong, especially given the ill-defined nature of its tertiary interactions. The kinetic parameters describing the kD5-dependent reaction are similar in this three-part system to those in a related two-part system that included D2 and D3 and covalent attachment of 5'-exon to D1 (Pyle, 1994). However, the 2X higher [KCl] used here stimulates the rate in both systems by 2-3-fold.

The ribozyme construct described here is relevant to splicing because a hydrolytic first step does not prevent the second step from occurring (Jarrell, 1988; and van der Veen, 1987). Furthermore, the first step is believed to be rate-limiting in two-step splicing by the ai5g group II intron and the rate of reaction by our ribozyme is comparable to that of overall splicing under similar reaction conditions (Jarrell, 1988; Pyle, 1994; Jacquier, 1987; Peebles, 1987; Chanfreau, 1994; and Peebles, 1993). The results observed here are not unique to the KCl conditions known to promote 5'-splice site hydrolysis. Efficient trans-cleavage of the small substrate occurs if KCl is replaced with any of other monovalent salts previously reported to promote the full splicing reaction.

Our results demonstrate that a group II ribozyme preferentially cleaves an rp diastereomer at the 5'-splice site in a reaction that appears to be analogous to the first step of splicing. There are several possible interpretations of this result, given that there are two distinct D5-catalyzed reactions that involve formation of the EBS-IBS pairings (Suchy, 1991). One possibility is that our system mimics the forward first step of splicing and that group II introns generally proceed with Rp stereospecificity during the first step. If this is so, it

is remarkable given that spliceosomal processing (for which group II introns are commonly invoked as models) proceeds with Sp stereospecificity during the first step (Moore, 1993; Maschhoff, 1993). It has been proposed that Rp stereospecificity, in either step of group II splicing, would indicate that group II introns are unrelated to the spliceosomal apparatus (Lamond, 1993). However, it is possible that hydrolysis reactions or reactions where the substrate is presented in-trans may result in different stereospecificity than that of cis-splicing. This might occur if the $\epsilon$-$\epsilon'$ interaction (between the +3,+4 and +116,+117 nts of ai5g) cannot form properly in-trans or if branch formation requires participation of additional subunits not present in our construct. In any case, we have characterized a minimal first step reaction which proceeds with comparable rates and subunit binding constants to those of more complicated systems and is specific for cleavage of sequences immediately surrounding the 5'-splice site. Thus, if stereospecificity of our system turns out to be different than that in cis-splicing, it means that we can build on our system to identify additional structures solely responsible for Sp branch formation. It would remain relevant to pre-mRNA splicing because snRNAs are believed to have evolved from group II domains that became separated and started to function in-trans, much like our system (Sharp, 1985; Cech, 1986). It is therefore important to understand how the group II intron can be faithfully divided into trans-ribozyme constructs.

A second possibility is that our system mimics a reversal of the second step of splicing and that the reaction may not proceed through the same active site as that utilized for branch formation. There are many indications that a reaction equivalent to the reverse of the second step of group II splicing can result in hydrolysis of linkages that follow 5'-exon sequences, particularly when certain structural features are lacking or under alternative

reaction conditions (Jarrell, 1988; Suchy, 1991). However, there are several indications that our reaction is not a reversal of the group II intron second step. D3 has previously been shown to be important for spliced exon-reopening, independent of its role in efficiency of the first step (Koch, 1992; Bachl, 1990). Our construct does not contain D3. Perhaps most importantly, our ribozyme does not readily cleave an oligonucleotide composed of ligated exons (nts -19 to -1 of 5'exon followed by the first 9 nts of the 3'-exon, data not shown). When provided in excess, this oligonucleotide is significantly less inhibitory to our reaction than an equal amount of 5'-splice-site substrate, thus implicating effects on both binding and chemistry. Therefore, the intronic sequences of our normal substrate are important for catalysis (perhaps verifying the importance of the $\epsilon$-$\epsilon'$ interaction) and our reaction mimics the forward first step in both sequence specificity and kinetic profile.

In addition, like the first step of splicing this reaction is inhibited by the presence of a C residue immediately 3' to cleavage site.

**Relationship to group II intron self-splicing:**

The ribozyme described here behaves in a manner consistent with the first step of group II self-splicing and therefore serves as a representative model for first step processes (Table 2). The fact that the reaction proceeds hydrolytically, rather than through branch-point attack, does not exclude the reaction from representing the first step for several reasons. The first step of splicing can proceed by hydrolysis without preventing the second step from occurring (van der Veen et al., 1987; Jarrell et al., 1988(b)). Therefore, the hydrolytic first step is a significant catalytic strategy that group II introns can exploit in order to effect splicing. Spontaneous attack of water at specific internucleotide linkages is not a

favorable process in the absence of a catalyst. The uncatalyzed rate for hydrolytic cleavage of the phosphodiester linkage has been estimated to be $3 \times 10^{-9}$ $min^{-1}$ (Herschlag and Cech. 1990). Given the kinetic parameters reported here, the group II intron provides a $1 \times 10^{7}$-fold rate-enhancement for hydrolytic cleavage at the 5'-splice site. Like RNase P, the group II intron contains active site functionalities that specifically promote this reaction. Recent studies of in-vitro self-splicing show that, under all reaction conditions, the first step continuously proceeds through both competing hydrolytic and transesterification pathways. It is well documented that the first step of self-splicing is rate-limiting in ai5g group II intron (Jacquier and Michel, 1987; Peebles et al., 1987; Chanfreau and Jacquier, 1993). Therefore, concrete kinetic parallels link the ribozyme reported here and the overall rate of group II intron self-splicing.

The reaction reported in this study might be considered analogous to the "spliced-exon-reopening" or SER reaction promoted by group II introns in high concentrations of KCl. This is unlikely given that ligated exons are a poor substrate for the ribozyme and that our results are not unique to the KCl conditions exclusively thought to promote SER (Jarrell et al., 1988(b). Efficient trans-cleavage of the small substrate occurs if KCl is replaced with any of other monovalent salts previously reported to promote the full splicing reaction. Another distinguishing feature is that the SER reaction is unaffected by C substitution at the +1 nucleotide (Peebles et al., 1993). The ribozyme reaction described here, like the first step of the full splicing reaction, is inhibited by the presence of a +1C, while any other nucleotide can function at that position. Although the reaction reported here is clearly distinct from SER, we consider SER to be a reaction related to the first step of group II intron splicing. If a group II intron does not rapidly release ligated exon products, they may be

susceptible to another nucleophilic attack at the splice junction. A common misconception is that the SER reaction is a reversal of the second step of splicing. Strictly speaking, this is impossible given that the second step of group II intron splicing occurs exclusively (and quite rapidly) through a transesterification in which the leaving group is the 3'-hydroxyl of the lariat or linear intron (Jarrell et al., 1988(b). Water is not the leaving group in the forward reaction, so water cannot be the attacking nucleophile in a reaction that is truly the reverse.

**Table 2: Kinetic effects of base changes at the ribozyme cleavage site**

| Substrate¥ | $k_{obs}$ (min-1)§ | relative rate |
|---|---|---|
| +1G (WT) | 0.036† | 1.0 |
| +1A | 0.053† | 1.5 |
| +1U | 0.028† | 0.79 |
| +1C | 0.0019‡ | 0.05 |
| ligated exons | 0.0068‡ | 0.19 |

Table 2 Legend:

¥ The sequence of transcribed substrates used to monitor mutational effect at the 5'-splice site is 5'-GGUGUGGUGGGACAUUUUC^NAGCGGUU, (Seq. ID No. 6) where N marks the +1 nucleotide at the cleavage site. The sequence of the ligated exons substrate is 5'-GGAGUGGUGGGACAUUUUC^ACUAUGUA, (Seq. ID No. 7) where the underlined nucleotides are 3'-exon sequences. Rates are not affected by identity of the first three 5'-nucleotides.

§ All rates were obtained simultaneously so it is exptectd that their relative error is very low (< 5%). The absolute error for $k_{obs}$ values reported in this study is ± 19% with 95% confidence (see legend, Table 1).

† Pseudo-first order plots were linear for at least four

reaction half-times (as in Fig. 9) and empirically determined endpoints of 92-94% were used for data correction.

† For these slower substrates, <20% reaction extent, using a 92% endpoint was used in calculation of $k_{obs}$.

## Example 2: Branch-Point Attack in group II introns is a highly reversible transesterifiction, providing a potential proofreading mechanism for 5'-splice site selection

### Summary

By examining the first step of group II intron splicing in the absence of the second step, we have found that there is an interplay of three distinct reactions at the 5'-splice site: branching, reverse-branching and hydrolytic cleavage. This approach has yielded the first kinetic parameters describing eukaryotic branching and establishes that group II intron catalysis can proceed on a rapid timescale. The efficient reversibility of the first step is due to increased conformational organization in the branched intermediate and it has several important mechanistic implications. Reversibility in the first step requires that the second step of splicing serve as a kinetic trap, thus driving splicing to completion and coordinating the first and second step of splicing. Facile reverse-branching also provides the intron with a proof-reading mechanism to control the fidelity of 5'-splice site selection and it provides a kinetic basis for the apparent mobility of group II introns.

### Introduction

Splicing of group II introns is essential for the expression of certain organellar genes in plants, yeast, and fungi (Michel et al., 1989). In the absence of proteins, many group II introns catalyze their own excision from primary transcripts, concurrent with the ligation of flanking exons. This process is initiated by a distinctive reaction

involving the nucleophilic attack of an intronic 2'-hydroxyl group on the phosphate of the 5'-splice site (Figure 10A), resulting in formation of a 2'-5' link between the first nucleotide of the intron and a bulged adenosine near the 3'-splice site (Peebles et al., 1986; Schmelzer and Schweyen, 1986; van der Veen et al., 1986). A bulged adenosine nucleophile, together with the ultimate release of intron as a lariat species are mechanistic features that are also observed during nuclear pre-mRNA splicing by the eukaryotic spliceosome (Moore et al., 1993). Homologies in mechanism (Peebles et al., 1986; Padgett et al., 1994), splice site selection, and morphology of particular RNA domains (Madhani and Guthrie, 1992) have contributed to the hypothesis that group II introns are evolutionary predecessors of spliceosomal small nuclear RNAs (snRNAs; U1, U2, U4, U5 and U6) and that snRNAs, rather than spliceosomal proteins, might compose the catalytic active sites of the spliceosome (Sharp, 1985; Cech, 1986; Jacquier, 1990). Specifically, it has been proposed that individual snRNA molecules evolved from distinct group II intron domains that became separated and began to function in-trans. An experimental basis for this idea is provided by the fact that group II intron domains can be transcribed as separate molecules and reassembled in-trans to catalyze splicing reactions (Jarrell et al., 1988a; Wissinger et al., 1992; Franzen et al., 1993; Pyle and Green, 1994; Michels and Pyle, 1995). It is likely that group II intron self-splicing is most closely related to the spliceosomal mechanism during the first step, when the branch-point 2'-hydroxyl group attacks the 5'-splice site. Therefore, because of its relationship to the spliceosome and its innate importance for gene expression in a wide variety of organisms, we examined the kinetic pathway for group II intron branch-point attack.

Group II introns have a conserved secondary structure that is organized into six domains arranged in the configuration of a wheel (Figure 10A, (Michel and Dujon, 1983) ).

Although a variety of substructures are evident in most group II introns, only a few of the domains are essential for group II intron reactivity in-vitro. Domains 1 and 5 (D1 and D5) are required for all reactions of the intron (Michel et al., 1989). They are sufficient for hydrolytic cleavage at the 5'-splice site (Koch et al., 1992; Franzen et al., 1993; Pyle and Green, 1994) and they catalyze the cleavage of small oligonucleotides (Michels and Pyle, 1995). In the latter reactions, D5 can be provided as a separate molecule that catalyzes hydrolysis in-trans. D5 is highly conserved and contains many of the nucleotides that are essential for catalytic activity by the intron. Domain 6 (D6) contains the highly conserved bulged adenosine which serves as the nucleophile during branch-mediated attack on the 5'-splice site. A small RNA containing only D5, D6 and a 3'-exon has been observed to attack the group II intron 5'-splice site and branch in-trans to form a Y-shaped molecule rather than a lariat (Jarrell et al., 1988a; Dib-Hajj et al., 1993). These trans-D56 reactions have been used to study trans-splicing, where branch formation is followed by exon ligation.

There are two steps involved in group II intron self-splicing: branch point attack on the 5'-splice site (step 1), followed by exon ligation (step 2) (Figure 10A). Intermediate products that arise after the first step of splicing are barely detectable unless mutations disrupt nucleotides important for the second step (Jacquier and Michel, 1987; Peebles et al., 1987). Therefore, the first step is considered to be the rate-limiting, or slow step of group II intron self-splicing (Jacquier and Jacquesson-Breuleux, 1991). Given that the second step of splicing is so fast and that the first step is rate-limiting, it would seem that the second step should not interfere with a kinetic analysis of the first step of splicing. However, the rapid second step may obscure features controlling the internal equilibrium of the first step and prevent proper

interpretation of specific kinetic parameters. Therefore, in order to characterize active-site functionalities and mechanistic features particular to the first step of splicing, we have examined the first step in isolation, using wild-type intron sequences. A similar rationale was used for the reconstruction of a group I intron from *Tetrahymena* into a ribozyme that reported solely on the transition-state for the first step of splicing (Zaug and Cech, 1986; Herschlag and Cech, 1990).

In order to obtain kinetic parameters for describing the mechanism of the first step, the experimental system was designed so that the reaction site and nucleophile were situated on two separate pieces of RNA. In this way, one can control the effective concentration of each RNA domain. By manipulating relative concentrations of the reaction components, one can drive formation of a saturated complex that undergoes a single reaction and thereby obtain kinetic parameters that reflect specific rate constants and units of free energy contributing to transition-state stabilization (Herschlag and Cech, 1990; Fedor and Uhlenbeck, 1992; Pyle and Green, 1994). This approach has the added advantage of minimizing effects of unfavorable conformations potentially introduced by intervening regions of RNA (e.g. Domains 2 and 4) that do not play a direct catalytic role in splicing (Kwakman et al., 1989) but have the ability to interfere if they lie between catalytic components in a cis-construct.

In this work we show that the first step of splicing is broken into several distinct mechanistic pathways: We find that transesterification (branching) is highly reversible and it competes at all times with a slower, irreversible hydrolytic pathway for 5'-splice site cleavage. By isolating the first step of splicing as a single process, we have identified its major mechanistic features, dissecting chemical from conformational steps. In so doing, we provide the first kinetic characterization of any branching pathway,

in which the 2'-hydroxyl group serves as the nucleophile in a transesterification reaction. The kinetic analyses show that group II-intron catalyzed reactions can proceed over the timescale of seconds, and thus we prove that they are inherently fast enough to function in biological processes. The pathways and reactivity revealed during this analysis raise important issues regarding the regulation of splicing by group II introns. Although group II introns can self-splice in-vitro, they are probably aided by associated proteins in-vivo (Saldhana et al., 1993). The mechanism we have provided illustrates potential roles for proteins and RNA substructures in controlling the forward progress of RNA splicing by group II introns and, by analogy, the eukaryotic spliceosome.

RESULTS:

**Experimental design:** To examine the first step without interference from the second step of splicing, we designed a system for observing a single trans-branching reaction between two RNA molecules. One RNA contains the 5'-exon joined to Domain 1-3 of the intron (exD123, Figure 10A) (Jarrell et al., 1988a). This molecule contains the 5'-splice site, which includes the phosphodiester linkage that is cleaved during the first step of splicing. The second RNA consists of Domain 5 joined to Domain 6 (D56, Figure 10 A and B). The D56 molecule contains the nucleophile (the branch-point adenosine) and catalytic D5, but it is not joined to 3'-exon nucleotides that might participate in a second step (Figure 10B). Because there is no 3'-exon attached to D56, exD123 and D56 should undergo the first, but not the second step of splicing. While it is common for group II introns to be divided into two trans-splicing parts , the RNAs are usually joined by extensive base pairing in the region of domain 4 (Wissinger et al., 1992; Chanfreau and Jacquier, 1994). In contrast, the two RNAs presented here are expected to associate primarily through tertiary interactions, since base-pairing interactions between D56

and other portions of the intron have not been identified. Thus, the kinetic information derived from this study will be applicable to the elusive problem of group II intron tertiary structure. We previously determined that D5 alone associates strongly with D1 or D123 through a network of undefined tertiary interactions (Pyle and Green, 1994). A molecule of D56 (Figure 10B) was therefore expected to readily bind exD123 (Figure 10A) with the added capability that the 2'-hydroxyl of D6 could attack the 5'-splice site, joining D56 to D123 and releasing 5'-exon. This would be consistent with previous results showing that trans-splicing occurs between D56/3'-exon and exD123 (Jarrell et al., 1988a; Dib-Hajj et al., 1993). Because D5 alone catalyzes hydrolytic cleavage of exD123 at the 5'-splice site (Jarrell et al., 1988a; Franzen et al., 1993; Pyle and Green, 1994), a D56 mutant (D56A-) was constructed in which the branch-point adenosine was deleted. This molecule would be expected to cleave the 5'-splice site through a hydrolytic reaction, but should be incapable of branching. By studying D56 simultaneously with D56A-, it was possible to quantitatively describe both hydrolytic and transesterification pathways of 5'-splice site attack without interference from the second step of splicing.

**The 5'-splice site undergoes multiple, competing reactions:** It was necessary to establish that branching takes place between D56 and exD123 and, given that there are several possible reactions that can take place at the 5'-splice site, the experiment was designed to visualize all products of D56 and D56A- attack on exD123. To this end, internally $^{32}$P-labeled exD123 in trace (1 nM) was combined with an excess of D56 or D56A- (3 $\mu$M). Based on previous estimations of D5 affinity (Pyle and Green, 1994), 3 $\mu$M is expected to be a saturating concentration. Under these single-turnover conditions, every $^{32}$P-exD123 molecule should be bound to a molecule of D56 (or D56A-) and reacting at its maximum rate. This single-turnover experimental design is

analogous to that in which $^{32}$P-exD123 was treated as substrate, while D5 was added in excess as the enzyme for hydrolytic cleavage of the 5'-splice site (Pyle and Green, 1994).

Time-courses of the reaction reveal that D56 forms a high-molecular weight branch, becoming attached to D123 (note bold band D56/D123, Figure 11A). At the same time, the previously observed 5'-splice site hydrolysis reaction competes with branching (bands D123 and D12, Figure 11A) (Jarrell et al., 1988a). During both the branching and hydrolysis reactions, 5'-exon is liberated (Figure 11A). The quantity of hydrolytic cleavage at the 5'-splice site is the same for D56 as it was for D56A-, but the latter molecule is unable to form a branch with exD123 (compare D56 and D56A- panels, Figure 11A). This establishes that a bulged adenosine in D6 is essential for 2'-hydroxyl group attack, as previously observed for full-splicing reactions (van der Veen et al., 1987). The lack of branched product for D56A- suggests that 2'-hydroxyl groups elsewhere on D6 do not attack the 5'-splice site and that the fidelity of branch-site selection is not compromised in our system.

Once it was apparent that D56 and exD123 underwent transesterification, it was important to establish reaction conditions that were appropriate for careful analysis of this reaction. A survey of monovalent and divalent ion effects gave results consistent with other studies indicating that high concentrations of ammonium salts (0.5 to 1.5 M) stimulate branching relative to hydrolytic pathways for 5'-splice site cleavage (data not shown) (Peebles et al., 1987; Jarrell et al., 1988b). Ammonium sulfate readily stimulated branching, but only after long and irreproducible lags in reactivity. Ammonium chloride also promoted branching, but without the concomitant lags and poor kinetic behavior associated with ammonium sulfate. Although different divalent ions, including $Ca^{2+}$, $Mg^{2+}$ and $Mn^{2+}$

(100 mM) were examined, only $Mg^{2+}$ supported activity. Provided that at least 50% of the divalent ion was $Mg^{2+}$, $Mn^{2+}$ could substitute for some of the divalent ion requirement. Based on these findings, the buffer salts chosen for this study were 500 mM $NH_4Cl$ and 100 mM $MgCl_2$, which are the lowest salt concentrations capable of promoting efficient branching. Temperature was also varied and 45°C was found to be the optimal condition for this reaction.

When a single-turnover ribozyme reaction is highly pH-dependent, it is likely to be limited by the rate of chemistry (Dahm et al., 1993; Pyle and Green, 1994; Zaug et al., 1994). Conversely, a lack of pH dependence under these conditions suggests that reaction rate is limited by conformational change. The D5-dependent hydrolysis of the 5'-splice site was previously shown to be highly pH dependent and limited in rate by the chemical step of reaction (Pyle and Green, 1994). Although it contains a relatively intact D6 duplex, D56A- reacted in an almost identical manner to that previously reported for D5 (Figure 11A) (Pyle and Green, 1994). D56 and D56A- hydrolysis reactions of the 5'-splice site are both highly pH-dependent and, by analogy to D5 reactions, they are probably limited by the rate of chemistry. The maximum saturating rate of hydrolysis by both D56 and D56A- at pH 7.0 under standard reaction conditions ($k_{cat}$ for hydrolysis) is 0.018 ± 0.0066 $min^{-1}$ (determined as described for $k_{obs}$ values in the legend to Figure 13, inset). However, the branching reaction promoted by D56 is not pH dependent from pH 5.5 to 8.0 (Figure 11A and data not shown). This indicates that the rate-limiting step for branching is something other than chemistry.

In addition to having different rate-limiting steps, the two competing reactions promoted by D56 have markedly different profiles of product evolution. The hydrolysis reaction proceeds slowly and exponentially, as expected for a pseudo-first order reaction similar to that catalyzed by D5 alone

(Figure 11B). Conversely, the branching reaction is very rapid in the first few minutes, resulting in a burst of activity that halts and then decreases before all the exD123 precursor is consumed (Figure 11C). This effect is particularly pronounced at pH values greater than or equal to 7.5, at which point hydrolysis of the 5'-splice site dominates over branching (Figure 11A). Late in reaction, the observed decrease in branched molecules could be due to instability of the 2'-5' linkage, causing it to break and produce linear D123. However, this is ruled out by identical kinetics of hydrolysis for the D56A- and D56 reactions (Figure 11A). If breakage accounted for the decline in branched species, D56-mediated hydrolysis would appear to go faster, because it would possess an additional pathway for generating the D123 product. The plateau observed for D56 branching could be due to a limited population of conformations that are competent for transesterification. Alternatively, the observed behavior could be due to a branching reaction that is reversible while hydrolysis is not. If transesterification were reversible, the population of branch would plateau early and, in the presence of an irreversible competing reaction, it would appear to eventually decline and lose ground to an irreversible process that proceeds to completion. By this model, the hydrolysis reaction behaves as a trap or sink, through which all of the material ultimately reacts. If the latter model is correct, there are three reactions going on at the 5'-splice site: branching, reverse branching and hydrolysis. One might be able to set up a complicated mathematical model describing this phenomenon and fit the data to it. This would be difficult given that one cannot obtain an individual reaction endpoint for each of the many competing reactions observed here. Instead, we decided it was better to simplify the experiment so that more straightforward kinetic analysis could be applied. Our approach involved redesigning the experiment so that branching could be examined without interference from the

competing hydrolysis reaction.

**Branching is studied in the absence of hydrolysis:** One way to study branching independent of any other reaction is to follow the conversion of $^{32}$P-D56 into branched species ($^{32}$P-D56/D123). Because the nucleophile during branching is the bulged adenosine 2'-hydroxyl on D56, D56 is changed in the course of reaction, becoming incorporated into a higher-molecular weight molecule. Although the hydrolysis reaction is occuring, there are several reasons why it should not influence the kinetic parameters of branching determined in this experiment. Provided that exD123 is unlabeled and in excess, the hydrolysis reaction will be undetectable because D56 catalyzes that reaction in a multiple-turnover manner, without becoming incorporated into the products. The nucleophile in the hydrolysis reaction is water or metal-coordinated hydroxide, neither of which is covalently connected to D56. The noncovalent binding properties of D56 are similar to those of D5 (see following sections) and, by analogy to D5, the on-and off-rates of D56 from exD123 are expected to be fast (Pyle and Green, 1994; Michels and Pyle, 1995). Therefore, trace quantities of $^{32}$P-D56 (1nM) and excess exD123 (3 $\mu$M) should react in a manner purely descriptive of the branching reaction. As anticipated, under these reaction conditions, the only process observed is covalent attachment of $^{32}$P-D56 to D123 (Figure 12A). In a denaturing gel, the product of that reaction comigrates with the branched species observed when $^{32}$P-exD123 is labeled and D56 is in excess.

The branched D56/D123 molecule was physically characterized in two ways: P1 nuclease was used to confirm the presence of a pA(2'-pG)3'-pU branched trinucleotide and the position of the branch was confirmed by sequencing. To characterize the branched trinucleotide, D56 was transcribed with $^{32}$P-UTP and then reacted with exD123. The resulting $^{32}$P-D56/D123 branched product was isolated and digested to completion

with P1 nuclease, which cannot cleave a branched trinucleotide (Reilly et al., 1989). Products of this reaction were then subjected to standard electrophoretic techniques which revealed the presence of an abundant branched trinucleotide (Figure 12C) (Moore and Sharp, 1993), formed through the proper combination of nucleotides. The sequencing experiments, which were performed using phosphorothioate interference (Gish and Eckstein, 1988; Schatz et al., 1991), resulted in ladders that terminated at the bulged A branch-point, rather than at the 3'-end of D6.

**Branching cannot proceed to completion:** Having simplified the transesterification (branching) reaction to a single conversion of unbranched to branched species, and having characterized the products of this transformation, it was then possible to derive kinetic parameters describing the reaction. The conversion of $^{32}$P-D56 to $^{32}$P-D56/D123 was monitored as a function of time (Figure 12A). Quantitation of the data revealed that, although the initial reaction is rapid and follows a single exponential, it proceeds to only ~50% completion (Figure 12B). Again, there are at least two explanations for this observation: Either only 50% of the D56 molecules are catalytically active, or the reaction is reversible. A hallmark of reversible reactions is that they will proceed to an apparent endpoint of less than 100% completion because a steady-state equilibrium is established between starting materials and products (Jencks, 1987). Experiments were therefore designed to test for this possibility and to measure the rate of any reverse reaction. Approaches for examining the reverse reaction are described. However, before proceeding with a discussion of the reverse reaction, it is important to point out mechanistic information that was obtained from timecourses of the forward reaction (Figure 12A, B).

**Branching kinetics reflect D56 tertiary interactions:** Rates of the forward branching reaction were obtained by

correcting the data for an observed 50% reaction endpoint (Figure 12B) and then plotting the evolution of product in a semilog format (Figure 13B). The data fell on a straight line and observed rates of reaction ($k_{obs}$) at a given [exD123] were taken from the slope of plots such as the one shown. The fact that the data are linear through four half-lives of reaction indicates that the reaction is "pseudo-first order", meaning that the $^{32}$P-D56 and exD123 react as if they were one molecule, in a complex. Linearity in this type of plot also indicates that the observed forward reaction is kinetically "well-behaved", representing a single process that is uniform from early to very late in the reaction. As shown in Figure 11, the rate of this reaction is not pH dependent, indicating that it is not limited in rate by the chemical step of reaction. In addition, the linear kinetic behavior takes place after a tiny burst of initial reaction, representing 4% of the total population of molecules (y-intercept, Figure 13B). Taken together, these data indicate that the rate-limiting step in the forward reaction for most of the molecules is not chemistry and is likely to represent the rate of conformational change of the D56/exD123 complex into a catalytically active state.

Although the forward rate of branching ($k_{obs}$) does not represent chemistry, the single-turnover design of this experiment and pseudo-first order kinetic behavior allowed us to use $k_{obs}$ values for estimating the relative affinity between D56 and exD123. This was accomplished by determining the rate of branching ($k_{obs}$) at a variety of [exD123]. The $k_{obs}$ values obtained from a collection of semilog plots (such as that shown in Figure 13B) were plotted as a function of [exD123]. As previously observed for D5-mediated reactions (Pyle and Green, 1994), these plots show saturable reaction behavior and fit a standard bimolecular binding curve (Figure 13A). Kinetic parameters for the branching reaction were obtained from the fit of

this binding curve to the data. The $k_{cat}$, or maximum rate of the forward reaction at saturation is 0.041 ± 0.0040 $min^{-1}$ and $K_m$ is 510 ± 160 nM. This $K_m$ for D56, as revealed through single-turnover branching reactions is remarkably similar to that of D5 (270 ± 25 nM), determined through single turnover kinetics, multiple turnover kinetics and direct binding assays (Pyle and Green, 1994). D56A- shows similar binding parameters, revealed through kinetic analysis of the hydrolysis reaction. These data suggest that D6 does not contribute additional tertiary interactions for stabilization of the group II active site and that most of the inter-domain binding interactions are contributed by D5 for both branching and hydrolysis reactions.

**Reverse-branching requires a 5'-exon nucleophile:** The reverse reaction is expected to occur when the 3'-hydroxyl group of the 5'-exon acts as a nucleophile, attacking the phosphate at the branch-site and releasing the 2'-hydroxyl of D56 as the leaving group. The first experiments conducted using internally labeled exD123 suggested that, because branched material decreased after an initial plateau, the branching reaction was fully reversible (Figure 11C). This was supported by the 50% endpoint observed in reactions of $^{32}$P-D56 with exD123 (Figure 12B). It has been proposed that the 5'-exon is not released immediately following attack at the 5'-splice site, but rather that it remains noncovalently attached to the intron (Jacquier and Rosbash, 1986). In the bound configuration, the associated exon can either perform the second step or, as observed here, it can attack the branch in a reverse of the first step. However, if the exon were removed and purified, it might be expected to have difficulty re-entering the active site.

In the previous experiments (Figure 11C and 12B), the 5'-exon was the long, 293-nucleotide exon normally attached to D123. To study reverse-branching, we generated this exon

from a large-scale reaction. It was gel-purified and then added to a reaction containing the branched $^{32}$P-D56/D123 material. However, after isolation and gel-purification, the 293- nucleotide 5'-exon was unable to react. Without the context of the adjacent intron, or perhaps as a consequence of purification, the long exon could not function as a nucleophile, possibly having adopted an unfavorable conformation that occluded its 3'-end. However, a short 5'-exon that contains all sequences necessary for interaction with the intron readily stimulated reverse branching after addition to $^{32}$P-D56/D123. This RNA, which was used in all subsequent kinetic analyses of reverse branching was 17-nucleotides in length, consisting of intact IBS1 and IBS2 sequences and terminating with the 3'-hydroxyl group of -1C (the last nucleotide of the 5'-exon). The sequence of this oligonucleotide is 5'-CGUGGUGGGACAUUUUC-3' (SEQ I.D. No. ___). Reactions of ai5g and other introns have been observed in the presence of short 5'-exons containing only the IBS sequences (Figure 10A) (Suchy and Schmelzer, 1991; Michels and Pyle, 1995). This is made possible by the fact that EBS-IBS interactions between exon and intron are very strong, with $K_d$ determined to be ~ 10 nM in the case of the ai5g intron (Michels and Pyle, 1995).

For studies of reverse-branching, $^{32}$P-D56/D123 was isolated from a denaturing polyacrylamide gel and purified according to standard techniques. Using the same reaction buffers and incubation temperature as the forward reaction, trace quantities of branched material (1 nM) were reacted with an excess of 5'-exon ( Figure 14A). During the ensuing reaction, the branched material rapidly gave rise to free $^{32}$P-D56 in a manner that reflected the strength of 5'-exon association. Timecourses of the reverse reaction were conducted in the presence of varying concentrations of 5'-exon. Consistent with strong interactions IBS-EBS reactions, the efficiency of reverse-branching (see below) decreased only at concentrations of 5'-exon lower than 10 nM, which serves as a rough estimate of $K_d$.

In the absence of 5'-exon, the branched species slowly hydrolyzed at the 2'-5' linkage (Figure 14A), but this process was extremely inefficient. Although addition of 5'-exon to $^{32}$P-D56/D123 stimulates efficient release of $^{32}$P-D56, this same apparent behavior would be observed if 5'-exon served only to enhance folding of D123, facilitating hydrolytic cleavage of the 2'-5'-linkage. Before reverse-branching could be conclusively established, it was necessary to observe covalent attachment of 5'-exon onto D123. This would prove that a transesterification, rather than a hydrolysis reaction, had occurred. To this end, 5'-exon was labeled with $^{32}$P and incubated with the branched species. As expected, upon reaction with $^{32}$P-D56-D123, the $^{32}$P-5'-exon shifts to the appropriate higher molecular weight upon reaction (727 nts), with concomitant appearance of proportional quantities of $^{32}$P-D56 (Figure 14B). Together, these results establish that the D56/D123 branch covalently reacts with 5-exon, releasing the 2'-hydroxyl of the bulged adenosine as the leaving group.

**The reverse-branching reaction is rapid:** Timecourses of reverse-branching reveal a rapid burst behavior indicative of a large population of highly reactive molecules. The kinetic data fall in two distinct phases, and rates of reverse branching were calculated by fitting the raw data to a biexponential equation (Figure 14C). This type of analysis allows one to obtain the rates of two different, simultaneous reactions. The burst phase of reaction, representing 19% of the molecules, proceeded with a rate of 1 $min^{-1}$. The remaining population of molecules reacted with a rate of 0.01 $min^{-1}$. For both phases of reaction, $^{32}$P-exon shift experiments analogous to those shown in Figure 14B confirmed that free D56 evolution was the result of 5'-exon attack, and not slow hydrolysis of the branched linkage. The fast phase of reaction ($k_{obs}$ = 1 $min^{-1}$) represents a lower limit for our estimate of the chemical rate for reverse branching. By contrast, the second slow phase of reaction,

like the forward reaction, is likely to represent the rate of conformational change from an inactive state to a state that is capable of reverse branching. It is highly significant that a higher fraction of molecules is conformationally predisposed to burst and react quickly in the reverse reaction (19%) in comparison to the forward reaction (4%). This suggests that the branched molecules are more preorganized and that the 2'-5' linkage, which brings the 5'-splice site close to catalytic functionalities on Domain 5, helps more of the population to assume a conformation capable of efficient reaction. These data confirm that, as indicated by behavior of the forward reaction, the reverse reaction is fast and can compete effectively during the first step of splicing (Figure 15). Were it not for a highly efficient second step of splicing, which engages every free 5'-exon as soon as it forms, the forward reaction would be unable to proceed to completion through branching.

DISCUSSION

One of the most distinctive features of self-splicing by group II introns is that the first step of splicing takes place through branch-point attack, resulting in production of a lariat intron (Peebles et al., 1986; Schmelzer and Schweyen, 1986; van der Veen et al., 1986). Here we show that this reaction occurs simultaneously with two other reactions, all of which have important implications for the mechanism of productive group II self-splicing and intron mobility. We observe that the forward branching reaction cannot proceed to completion on its own. The reason for this is that a rapid reverse reaction becomes dominant once a population of branched molecules is established. The 5'-exon that is released during the forward branching reaction is a highly reactive nucleophile in the reverse reaction, readily debranching the 2'-5' linkage. Our data suggests that the reverse reaction is actually more facile than the forward reaction due to greater proximity of active-site components. Another reaction that competes during forward

branching is hydrolytic cleavage of the 5'-splice site. Although this route is slower than branching, it is irreversible and is observed to drive the first step to completion in the absence of other reactions (such as the second step of splicing).

**An alternative pathway for the first step of splicing:** The hydrolytic pathway makes a significant contribution to the forward reaction and to splicing in general. Although it is slow, it is limited by the chemical step and not by conformational heterogeneity of the intronic components (Pyle and Green, 1994). Catalyzed by D5, this reaction appears to serve as a default reaction that promotes proper 5'-splice site cleavage regardless of intronic conformational problems. The reaction appears to take place because of an inherently large strain on the scissile bond of the 5'-splice site, rendering that linkage susceptible to attack by nucleophiles such as water or $Mg^{2+}$-bound hydroxide. The hydrolysis reaction is irreversible, and might therefore be an important means by which genes ensure that forward splicing takes place without risking intron mobility into other parts of their genome. In-vivo, there may be proteins that facilitate the hydrolytic splicing pathway and others that suppress it in favor of branching. Proteins that may belong to the latter class have already been implicated in intron mobility, a process that would benefit from splicing through branching (Lambowitz, 1989; Lambowitz and Belfort, 1993).

**Conformational limitations on 5'-splice site transesterifications:** When monitored in the absence of hydrolysis, the kinetics of the forward branching reaction are remarkably linear for up to four half-times of reaction and they do not display lags when plotted in a semi-log format. This indicates that there is no set population that is particularly slow to fold and that each molecule has the capability to react through branching. The fact that the

observed rates (slopes of semilog reaction plots) do not represent the chemical step and that the reaction shows saturable behavior at 510 nM exD123 indicates that D56 readily binds exD123, but that in the bound configuration it samples a set of conformational states, only a subset of which are competent for branching. For the forward and reverse reactions, the population of D56/exD123 complexes that react immediately (4 and 19%, respectively) represent the fraction of molecules that started in the proper conformational state for reaction. Once this population is consumed, the remaining D56/exD123 molecules react more slowly, as if limited by the rate of a single conformational change. It is important to point out that this is probably not a directed, single conformational change from one state to another. Macromolecules typically sample many conformations in a random fashion. The slow rates of transesterification following bursts of activity may therefore represent the average time of stochastic sampling before the complex (either $D56_{\delta}exD123$ in the forward reaction or 5'-$exon_{\delta}D56/D123$ in the reverse) reaches a conformation that is kinetically competent to react.

It is interesting that the $K_m$ for D56 (510 nM), as reflected through the kinetics of the branching reaction (Figure 13), is similar to that of D5 (270 nM), as reflected through direct binding assays and catalysis of the hydrolysis reaction (Pyle and Green, 1994). This indicates that D6 does not contribute additional tertiary interactions beyond those already provided by D5. Our data support the notion that D5 functions to dock into the core and position D6 correctly such that the bulged adenosine is in position to attack the 5'-splice site (Dib-Hajj et al., 1993). Further evidence for the idea that D5 behaves similarly in the presence and absence of D6 comes from rates of hydrolytic cleavage at the 5'-splice site. Just as D5 functions as a true enzyme for exD123 hydrolysis (Franzen et al., 1993; Pyle and Green, 1994), so does D56. The rates of hydrolysis

catalyzed by D56 and D56A- (0.018 $min^{-1}$) are only slightly slower than that of D5 alone (0.041 $min^{-1}$) (Figure 15). The small difference in their hydrolysis rates may suggest that D6 shields the 5'-splice site from attack of water, or that it sterically occludes the approach of solvent-based nucleophiles, but such effects are probably small given the minor differences in hydrolytic rate between D5 and D56 analogs.

**Rapid transesterification by a group II intron:** It is important to consider why the branching reaction proceeds so readily in reverse. After all, a 2'-hydroxyl group is a better nucleophile than a 3'-hydroxyl group (Sulston et al., 1968) so if nucleophilicity were the only consideration, the reaction would probably proceed only in the forward direction. However, the forward reaction creates a higher-energy species than the reverse reaction. The Y-shaped, or lariat species created after the first step places catalytically essential Domains 5 and 6, together with the 3'-splice site, in immediate proximity to intronic portions of the 5'-splice site. The high-energy nature of the branched species is also supported by our data, which indicate that the 2'-5' bond is weaker than 3'-5' linkages that might replace it (Figure 14A). It has been proposed that the constrained lariat structure formed after the first step of splicing drives the second step (Jacquier and Jacquesson-Breuleux, 1991). Here we show that this high-energy intermediate can also drive the first step backwards, so it is not necessarily a determinant in second-step efficiency. Conformational preorganization of the branched species is expected to be especially pronounced with intact group II introns because, in addition to covalent connection, a lariat intron circularly constrains active-site residues so that they are held in spatial proximity. The efficiency of reverse branching in this two-part system should not be attributed to the fact that catalytic domains are located on one molecule in the reverse reaction, while

they are divided on two molecules in the forward direction. This entropic effect is countered by the fact that all of the reactions reported here were studied at saturation. That is, even though D5 and D6 were on a separate molecule in the forward reaction, concentrations were manipulated such that they were fully bound to exD123. This was confirmed by the pseudo-first order (acting as one molecule instead of two) kinetic behavior of the complex. It is potentially harder to ensure docking of D56 when these domains are physically connected to D123 in-cis. This is because intervening RNA conformations can prevent an RNA molecule from assuming the docked state and it would be difficult to kinetically detect this effect.

The reverse reaction appears to be so rapid not because it is necessarily faster in the chemical step (although the lower limit for that step is a very rapid 1 $min^{-1}$), but because more of the molecules are conformationally competent to chemically react. Note that the burst in the forward reaction represents approximately 4% of the total population, while the burst in the reverse reaction is 19% of the population. In both forward and reverse reactions, bursts are followed by slower reaction phases ($k_{conf(fwd)}$ = 0.041 $min^{-1}$; $k_{conf(rev)}$ = 0.010 $min^{-1}$) that represent the rate of conformational change into catalytically competent forms (Figure 15). Thus, there is a direct catalytic advantage conferred by constraining the branched molecules. In-vivo there are probably proteins that facilitate folding into fully-active conformations for reverse or forward reactions, depending on whether the protein is designed to stimulate splicing or intron mobility (Lambowitz and Belfort, 1993).

In addition to providing a mechanistic picture of competing reactions during the first step of splicing, this study reveals for the first time that group II introns have the innate ability to catalyze fast reactions. For the most

active population of molecules, the reverse of the first step proceeds with rate of 1 $min^{-1}$, and judging from the small burst observed in the forward direction, its rate is at least that fast. A rate constant of this magnitude is observed for the action of many protein enzymes during other biologically significant reactions, such as DNA cleavage by restriction endonuclease Eco R1 (Walsh, 1979). It is also similar to the $k_{cat}$ values observed for other ribozymes (Pan et al., 1993). Previous studies indicate that the rates of group II intron-catalyzed reactions are relatively slow, with best half-times of ~ 10 minutes under ionic conditions similar to those used here (Michel et al., 1989). It is now clear that these apparently slower rates are due to constructs that react through hydrolytic pathways or that are limited in overall rate by unfavorable conformations. Without examining early sets of timepoints during transesterification, it would have been difficult to discern if these reactions were fast. In the absence of kinetic data demonstrating that a chemical step of RNA catalysis by group II introns could be efficient, there remained the possibility that intron-associated proteins were required for composition of the active site in-vivo. This study demonstrates that, given the proper conformational state, RNA structures within group II introns can catalyze reactions on a biologically feasible timescale.

**Relevance of a facile back-reaction during the first step of splicing:** While two-step splicing would obscure the apparent effects of branch reversion, the reaction is nonetheless important. In fact, a reversible branching reaction confers significant advantages. First, it provides a check for the fidelity of 5'-splice site cleavage. If an improper 5'-splice site were chosen during branch-point attack, the intron would lose interactions required for the second step of splicing, such as the interaction between first and penultimate nucleotides of the intron (Chanfreau and Jacquier, 1993; Peebles et al., 1993).

Facile branch reversion, particularly when mis-branching slowed the second step, would give the intron another opportunity to start over again and find the correct 5' splice site. This proof-reading mechanism would need to be verified by experiments showing that mis-cleaved 5'-exons can attack and reverse formation of a cryptic branched species. Reversibility might also prevent mis-splicing during transcription. Any branching reaction that takes place before important residues in the 3'-exon are free of template and polymerase would readily reverse. Thus, until the RNA is fully competent for an accurate second step, branching reversal will effectively prevent splicing. A similar mechanism may help control the fidelity of splice-site selection during spliceosomal processing. Even in that case, the formation of the 2'-5' link after the first step of splicing might create a higher-energy intermediate in which a reverse reaction is favorable. However, in that case the control of this reaction is probably mediated by snRNPs and associated factors rather than by intronic substructures. The spliceosomal protein Prp 16 has recently been implicated in the controlling the fidelity of branch-site selection (Burgess and Guthrie, 1993). Similarly, a protein may control fidelity of 5'-splice-site selection by exploiting the potential reversibility of spliceosomal branching reactions.

The second step of group II intron self-splicing is known to be a very rapid reaction that proceeds almost entirely to completion, while the first step of splicing is known to be rate-limiting (Jacquier and Jacquesson-Breuleux, 1991). Our results provide a clear reason for the rapidity of the second step: it serves as a kinetic trap for those molecules that succeed in branching. Before the first step of splicing can reverse, the second step rapidly ensues and the full splicing reaction is driven to completion. Thus, there is necessarily kinetic coordination of the first and second steps of splicing in group II introns. While

precision timing of snRNP binding and release helps to coordinate the two steps of spliceosomal processing, RNA substructures in the group II intron have adopted the thermodynamic balance and kinetic control necessary for driving a series of competing reactions to a common conclusion, resulting in fully spliced RNA.

One of the most intriguing consequences of efficient branching reversal is its potential role in intron mobility. Group II introns are infectious elements that have scattered widely through the genomes of many different organisms (Lambowitz, 1989). There are several models for the mobility of self-splicing introns, some of which do not even involve the splicing of intronic RNA (Lambowitz and Belfort, 1993). However, one model for intron mobility proposes that liberated lariat introns can bind other RNA molecules, recognizing sequences with a high degree of complementary to their own EBS regions. Through a series of two successive reverse transesterifications, the intron is believed to insert itself into the host RNA. Group II introns often contain open reading frames for proteins with reverse-transcriptase domains (Saldhana et al., 1993), so a DNA copy of the modified transcript might be made and incorporated during recombination into the host genome. This model for intron mobility rests on the observation that excised introns can reverse-splice into new RNA molecules, targeting sequences complementary to endogenous EBS sequences. Reverse splicing has been observed for both group I (Woodson and Cech, 1989) and group II introns (Augustin et al., 1990; Měrl and Schmelzer, 1990; Muller et al., 1991). This observation does not necessarily reflect the relative efficiency of forward and reverse reactions. After all, the first step of group I intron self-splicing does not proceed with sufficient efficiency for its products to affect the apparent endpoint of forward reactions. However, an important implication of reverse-splicing reactions is that the branching reaction of group II introns is capable of

reversal. The attack of 5'-exon on mutant lariats of intron bI1 was examined previously, and these studies suggested that the most difficult step in reverse-splicing was reversal of the branching reaction (Augustin et al., 1990). This may have been due to a lack of kinetic control on the system or the use of mutants rather than wild-type sequences. It is now well-established that many modifications that affect the second step of group II self-splicing also affect the first step (Chanfreau and Jacquier, 1994). Therefore, an important ramification of our analysis is that the final step necessary for reverse splicing, the reverse-branching reaction, is actually facile and that the intron can readily surmount this particular kinetic barrier to reintegration into an RNA site. Proteins that promote intron mobility, such as the maturases (Lambowitz and Belfort, 1993), may act by simply inhibiting the forward second step of splicing, thereby exploiting the natural propensity of the branching reaction to proceed in reverse.

Conclusions

Together, these data show that the first step of splicing involves the interplay of three reactions, all of which have the capability to serve important biological functions. The hydrolysis reaction results in proper splice-site selection, and although it is a slow chemically limited process, it is able to drive the first step of splicing to completion. The two transesterification reactions, branching and reverse branching, are chemically rapid and partially limited in rate by slow conformational change. The forward branching reaction provides kinetic parameters indicating that tertiary interactions with D5 mediate association of D56 in the active site. The forward reaction cannot go to completion because it is a highly reversible transformation. The reverse reaction is very fast, proceeding through a higher initial population of conformationally competent molecules. The efficiency of reverse branching would permit it to aid in proof-reading of 5'-splice-site

selection and may explain the facile mobility of group II introns. The presence of an efficient back-reaction during the first step of splicing necessitates a rapid second step in order to drive group II intron self-splicing to completion.

EXPERIMENTAL DESCRIPTION

**Plasmids, templates, and transcription of RNA:** The group II intron examined in this study is intron 5γ of the COX1 gene encoding the mitochondrial cytochrome oxidase subunit 1 of *Saccharomyces cerevisiae* (ai5g). Plasmids pJDI3'-673 and pJDI5'-75 were kindly provided by Dr. Philip S. Perlman (Jarrell et al., 1988a). Plasmid pT7-D56 was constructed by Kunkel mutagenesis of pJDI5'-75 with the following modification to published protocol (Kunkel et al., 1991): Annealing of the mutagenic oligonucleotide was done by heating to 65;C for 5 minutes and then placing samples directly on ice. Mutagenic oligonucleotides were synthesized on an ABI 392 DNA/RNA synthesizer and purified by PAGE. Plasmid pT7-D56 contains domains 5 and 6 with a unique EcoRV site at the 3' end of D6. Plasmid pT7-D56A- was similarly constructed except that the branch point adenosine has been deleted. Plasmid templates were linearized by restriction with EcoRV before transcription with T7 RNA polymerase (Davanloo et al., 1984). Transcription of pJDI3'-673 and pJDI5'-75 to make exD123 and D5 RNA, respectively, has already been described (Pyle and Green, 1994). Transcriptions of pT7-D56 and pT7-D56A- to make D56 and D56A- RNA, respectively, were done under identical conditions. Synthetic 5'-exon analog 5'-CGUGGUGGGACAUUUUC-3' (SEQ I.D. No.: ___) was synthesized from 2'-silyl protected RNA phosphoramidites on an ABI 392 DNA/RNA synthesizer. This RNA was purified using a reverse-phase OPC (ABI) column followed by PAGE.

**Forward reaction conditions:** Reactions were performed in 0.65 ml Slickseal tubes (National Scientific) in a final

volume of 10 or 20 μl. RNA provided in excess was unlabeled, while RNA provided in trace was labeled with $^{32}P$, as indicated. To begin a reaction, aliquots of exD123 and D56 RNA (stored in 10mM MOPS and 1mM EDTA pH 6.5) were individually diluted and combined with 1M MOPS pH 7.0 to a final concentration of 55 mM MOPS pH 7.0. The RNA components were then heated to 95¡C for 1 minute to unfold alternative structures formed by storage at -20¡C (Groebe and Uhlenbeck, 1988; Walstrum and Uhlenbeck, 1990). RNA samples were cooled at room temperature for 15 seconds and then briefly spun in a microcentrifuge. Reaction was initiated by simultaneously combining the exD123 sample, the D56 sample and a concentrated salt solution such that the final reaction contained both RNA components at indicated concentrations and a reaction buffer of 40 mM MOPS pH 7.0, 100 mM $MgCl_2$ and 500 mM $NH_4Cl$. Results were identical if reaction was initiated by combining D56 and exD123 RNAs that had been previously incubated individually with reaction buffer. After either initiation protocol, reactions were incubated at 45°C in a water bath, aliquots were removed at selected timepoints and combined with equal volume of gel loading buffer (10 M urea, 0.1% xylene cyanol and bromphenol blue dyes, 8.3% sucrose, 40 mM tris pH 7.5 and 0.83 mM EDTA, final concentrations) before placing on ice. Samples were then subjected to electrophoresis on denaturing 4% polyacrylamide gels, which were then dried and quantitated on a Packard Instant Imager using methods previously described (Pyle and Green, 1994). Experiments were performed to determine optimal reaction conditions (described above) for quantitation of branching kinetics: a) the pH was varied from 5.5 to 8.0, b) the temperature was varied from 37° to 50°C, and c) the salts tested were 100 mM $MgCl_2$ (99.99% pure, Aldrich), $MgSO_4$ or $MnCl_2$ combined with 500 mM $NH_4Cl$, $(NH_4)_2SO_4$, KCl or NaCl.

**Determination of kinetic parameters for the forward reaction:** The forward reactions fit a single exponential

with a plateau at 50% reaction completion. To quantitate the rates of these reactions, the data was corrected for a 50% reaction endpoint and the natural logarithm of the quantity (1- fraction product) was plotted vs. time in a semilog format as shown in the inset to Figure 13. The equation of the line was solved for the reaction half-time, which was converted to $k_{obs}$ (the $k_{obs} = 0.693 / t_{1/2}$). The same rate constant would also be obtained by fitting the raw data to a single exponential, but it would be harder to see deviations from reaction homogeneity at long reaction times. Initial bursts of reaction were estimated from the y-intercept of semilog plots. The rate of the forward hydrolysis reaction was determined by quantitating D56A-cleavage of exD123, using published procedures for studying D5-catalyzed reaction (Pyle and Green, 1994). The $K_m$ for the D56-exD123 complex and the overall $k_{cat}$ for the forward reaction were determined by plotting $k_{obs}$ values as a function of [exD123] (Figure 13). The data were fit using a published equation describing 1:1 bimolecular association (Pyle and Green, 1994). $K_m$ and $k_{cat}$ values, with accompanying standard errors were determined by best fit to the equation.

**Branch isolation:** Trace amounts of $^{32}P$-labeled D56 (10 nM, labeled internally or at the 5'-end) was incubated with excess exD123 (1$\mu$M) in 40 mM MOPS pH 7.0, 100mM $MgCl_2$, 500mM $NH_4Cl$ at 45¡C for 2-3 hours. After incubation, the sample was ethanol precipitated and then resuspended in 15$\mu$l MOPS storage buffer (10 mM MOPS, 1 mM EDTA, pH 6.5) and 15 $\mu$l gel loading buffer (described above) before electrophoresis on a 4% denaturing polyacrylamide gel. The gel was transferred to a piece of X-ray film that had been developed and then washed thoroughly. The gel was analyzed for 5-10 minutes on a Packard Instantimager and the actual-size image was laser-printed on a transparency sheet. By aligning the printout with the scanner coordinates of the gel, bands were easily identified and were excised with precision. Excised bands were eluted into MOPS storage buffer overnight before

ethanol precipitation and resuspension.

**Characterization of branched product:** In order to sequence the position of the branch site, phosphorothioate-doped D56 was transcribed using small amounts of specific α-thio A, C, U or G nucleotide triphosphates. The transcripts were $^{32}$P-end-labeled, reacted with exD123 to form the branched species, purified and then isolated. Iodine cleavage of phosphorothioates in the branched molecules, followed by electrophoresis on a high resolution polyacrylamide gel resulted in base-specific sequencing ladders terminating at the bulged A branch-point in D6. A large gap separated this terminal cleavage from the branched material at the top of the gel. For analysis of the branched trinucleotide, P1 nuclease was used to digest $^{32}$P-D56/D123 (typically 10,000 to 20,000 cpm branched RNA) in a buffer containing 25 mM sodium acetate pH 5.3, 0.025 μg/μl tRNA, and 0.005 units/μl P1 nuclease (Boehringer). Reactions were incubated for 5-15 min at 37¡C before adding to equal volumes of quench buffer (36% formamide, 0.1X TBE, 1.8% sucrose, 0.02% xylene cyanol dye and 25 mM EDTA final concentrations) and loading on a denaturing 20% polyacrylamide gel. Following the separation of mononucleotide (pU) from branched trinucleotide pA(2'-pG)3'-pU by electrophoresis, gels were dried and quantitated on a Packard Instantimager.

**Conditions for reverse reaction and kinetic analysis:** Stocks of branched D56/D123 species (labeled either at the 5'-end of D56 or internally in D56) and 5'-exon were diluted and reacted using the same procedures, buffers and reaction conditions described above for the forward reaction. The $k_{obs}$ values for the reverse reaction were obtained by fitting the fraction reacted vs. time to a biexponential equation:

$$y = Ae^{-k1t} + (1-A)e^{-k2t}$$

where A = the fraction of molecules that participates in the

fast phase of reaction, $k_1$ = rate of the fast phase (burst, possibly chemistry), $k_2$ = rate of the slow phase (conformational change) and t = time along the x-axis.

## Example 3: Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups

Summary:

Group II introns are self-splicing RNAs that catalyze a variety of important biochemical transformations. These autocatalytic molecules can be reconfigured into highly specific, multiple-turnover ribozymes that cleave small oligonucleotides in-trans. Unlike other catalytic RNA molecules, the group II ribozymes cleave DNA linkages almost as readily as RNA linkages. A ribozyme variant containing intron Domain 3 cleaves DNA linkages with an efficiency comparable to that of restriction endonuclease EcoRI. By probing the effects of single deoxynucleotides along the substrate strand, the ribozymes were found to bind substrate without engaging 2'-hydroxyl groups for ground-state stabilization. The ribose 2'-hydroxyl group at the cleavage site plays a minor role in transition-state stabilization by these ribozymes, suggesting a catalytic strategy that broadens the scope of potential substrates. Because they also do not engage 2'-hydroxyl groups in substrate binding, group II intron ribozymes may rely exclusively on base-pairing for molecular recognition.

### Introduction:

The ribose 2'-hydroxyl group is widely regarded as an important functionality for molecular recognition and chemical catalysis by RNA [Musier-Forsyth and Shimel, 1992; Pyle et al. 1992; Herschlag et al., 1993; Smith and Pace, 1993; and Stobel and Cech, 1993]. Evidence for its participation as either a hydrogen bond donor or acceptor demonstrate that the 2'-hydroxyl group is a multifunctional

recognition determinant [Pyle et al. 1992; Herschlag et al., 1993; Smith and Pace, 1993; Yap and Musier-Forsyth, 1995; Pley, H.M. et al., 1994; and Pley, H. M. et al., 1994]. For ribozymes in the hammerhead mechanistic class (class A), the 2'-hydroxyl group at the substrate cleavage site behaves as a nucleophile in the phosphodiester cleavage reaction [Pan et al., 1993]. For ribozymes in another mechanistic class (class B), which includes RNase P, group I introns and group II introns, the 2'-hydroxyl group adjacent to the scissile bond is not the attacking nucleophile (Scheme 1) [Pyle, 1993]. Nonetheless, in two examples (group I intron ribozymes and RNase P), this 2'-hydroxyl at the cleavage site has been observed to play an important role in transition-state stabilization [Smith and Pace, 1993, Herschlag et al., 1993; and Perreault and Altman, 1992]. Neighboring 2'-hydroxyl groups have also been observed to contribute to substrate recognition and binding [Pyle et al. 1992; Herschlag et al. 1993; Strobel and Cech, 1993 Pyle and Cech 1991; and Bevilacqua and Turner, 1991]. Although 2'-hydroxyls within a core region (Domain 5) of group II introns are essential for binding and catalysis, the specific role of 2'-hydroxyl groups on substrates for group II intron ribozymes has not been investigated.

Through the process of directed molecular evolution, a group I intron ribozyme was evolved to cleave DNA linkages with relatively high efficiency [Robertson and Joyce, 1990; and Tsang and Joyce, 1994]. But to date, there have been no examples of "natural ribozymes" that cleave DNA with an efficiency comparable to their cleavage of RNA. The term "natural" is used rather loosely here since most ribozymes are man-made pieces, comprising the actives site of self-cleaving RNA molecules. However, the term is used here to differentiate wild-type active-site sequences from those that have been optimized by the process of directed molecular evolution or selection. Regardless of their derivation, ribozymes that make little use of substrate 2'-

hydroxyl groups are an important source of information about alternative strategies in the catalytic repertoire of nucleic acid enzymes. In addition, this behavior by natural ribozymes may point to a biological impact of ribozymes on genomic DNA. The data presented here indicate that certain ribozymes derived from a group II intron stimulate efficient catalysis without engaging the 2'-hydroxyl groups on their target substrates.

Self-splicing group II introns have been found in the organellar genes of plants, lower eukaryotes and also in prokaryotes [Pyle, 1995; and Michel and Ferat, 1995]. The excision of ubiquitous group II introns is essential for metabolism in many of these of organisms. Group II introns can be arranged into six domains, of which domain 1 (D1) and domain 5 (D5) have been found to be essential for catalytic activity [Koch et al., 1992; and Michels and Pyle, 1995]. Other domains are dispensable, although certain deletions result in reduction of particular activities [Pyle, 1995]. Self-splicing can proceed via either one of two pathways. One pathway involves initial 5' splice site attack by the 2'-hydroxyl group from the bulged adenosine of domain 6 (D6), which generates a "lariat" intron that is ultimately released after the second step of splicing [Peebles et al., 1986; and Schmelzer and Schweyen, 1986; and van der Veen et al., 1986]. Alternatively, water can act as the nucleophile in the first step, resulting in ligated exons and release of linear intron [Jacquier and Rosbash, 1986; van der Veen et al., 1987; Jarrell et al., 1988b; and Daniels et al., 1995]. A recent study has demonstrated that hydrolytic attack is a legitimate and common pathway for in-vitro self-splicing of group II introns [Daniels et al., 1995].

In spite of their biological importance, there have been relatively few investigations of the structural features and mechanistic enzymology of group II introns [Jarrell et al.,

1988; Franzen et al., 1993; Pyle and Green, 1994; Padgett et al., 1994; and Podar et al., 1995]. Recently, a detailed kinetic study on a ribozyme derived from the group II intron ai5g was reported [Michels and Pyle, 1995]. That work established a minimal group II ribozyme system in which D1, acting as an enzyme along with D5 as cofactor, cleaves short RNA oligonucleotides analogous in sequence to the 5' splice junction (Scheme 2). Similar to hydrolytic splicing reactions (in either forward or reverse splicing), the nucleophile is water or hydroxide (Scheme 1). As expected for base pairing between the intron binding sites (IBS1 and 2) and exon binding sites (EBS1 and 2), D1 binds to its substrate strongly with a $K_d$ of 6.3 nM. The cleavage reaction is rate limited by chemistry, with $k_{cat} = k_{chem} = \sim 0.03$ $min^{-1}$ at saturation. Using this kinetic framework, it was possible to incorporate single-atom changes into the substrate and probe the reaction mechanism.

In this investigation, the role of 2' hydroxyl groups on the oligonucleotide substrate for a group II ribozyme was examined by incorporating deoxynucleotides into specific positions along the strand. By monitoring the effects of deoxynucleotide substituents on catalysis and binding to the ribozyme, it was possible to evaluate the role of these 2' hydroxyl groups. The data indicate that, for two different ribozyme constructs, 2' hydroxyl groups along the substrate backbone are minimally involved in tertiary contacts that stabilize binding to the ribozyme or stimulate catalysis during the chemical step. In one case, deoxynucleotide linkages are cleaved with the highest efficiency ($k_{cat}/K_m$) observed for any ribozyme to date, having a $k_{cat}$ comparable to that of certain protein restriction enzymes. Although significant effects are observed for deoxynucleotides at the cleavage site and immediately adjacent, these effects are small compared to those observed for other ribozymes in mechanistic class B, which show ~$10^3$-fold effects on $k_{chem}$ [Herschlag et al., 1993; Smith and Pace, 1993; and Herschlag

et al., 1993]. Trivial explanations for the effects reported here are ruled out by kinetic experiments that reveal, for both ribose and deoxyribose substrates, that the reactions proceed with the same rate-limiting step. These data suggest that, for certain types of reactions, group II intron ribozymes may utilize a novel catalytic strategy that bypasses nearby 2' hydroxyl groups and exploits other factors for transition-state stabilization.

**Experimental Procedures:**

*Preparation of RNA:* RNA transcriptions of D1 (425 nt) and D5 (58 nt), from plasmids pT7D1 and pJDI5'-75 respectively, were carried out as previously described [Michels and Pyle, 1995]. Ribozyme D123 (710 nts) was transcribed from plasmid pT7D123, which was cut with Hind III. This plasmid was constructed by Kunkel mutagenesis [Kunkel et al., 1991] from plasmid pT7I3'-675 [Jarrell et al., 1988; and Pyle and Greene, 1994]. Mutagenesis removed the 5'-exon from the parent plasmid, such that the first nucleotide of the resulting transcript corresponds to the first nucleotide of the ai5g group II intron from yeast mitochondria. Plasmids pJDI5'-75 and pJDI3'-675 were used. Oligonucleotide substrates, all ribose (rS) or with deoxyribonucleotides (chimeric substrates) selectively placed at specific positions, were synthesized on an ABI 392 DNA/RNA synthesizer, and worked up according to standard procedures [Scaringe et al., 1990]. Oligonucleotides were composed of the same base sequence which, for the all ribose substrate (rS) was 5'CGUGGUGGGACAUUUUC˅GAGCGGU (SEQ I.D. No. ___). Underlined positions represent IBS2 and IBS1 respectively, dashed underline represents the intronic sequences and ˅ indicates the cleavage site. The d(IBS1) block substrate was 5'CGUGGUGGGAdCdAdUdUdUdUC˅GAGCGGU, (SEQ I.D. No. ___) in which all IBS1 residues other than the nucleotide at the cleavage site were deoxynucleotides (bold). The d(IBS2) substrate contained 5'CdGdUdGdGdGdUdGGGACAUUUUC˅GAGCGGU (SEQ. ID No. ___). The d(link) substrate contained

deoxynucleotide substitutions at the three positions between IBS1 and IBS2 (as above). The d(intron) substrate contained the residues 5'CGUGGUGGGACAUUUUC˅dGdAdGdCdGdGdU. (SEQ I.D. No. ___) For substrates containing single deoxynucleotide substituents, they are named according to the specific position relative to former exon and intron sequences. For example, -1dC is 5'CGUGGUGGGACAUUUUdC˅GAGCGGU, (SEQ I.D. No. ___) and -2dU represents 5'CGUGGUGGGACAUUUdUC˅GAGCGGU, (SEQ I.D. No. ___) etc.

**Reaction conditions for kinetic analyses:** Characterization of preliminary reaction rates for the d(IBS2), d(link), d(IBS1) and d(intron) substrates was carried out under single turnover conditions with saturating ribozyme. The reaction buffer contained 1 M KCl, 100 mM $MgCl_2$, and 80 mM 3-[N-Morpholino]propanesulfonic acid (MOPS) pH 7.5. Reactions were usually carried out in 20 ul volumes at 42ûC, with excess ribozyme (100 nM - 250 nM D1 or D123, 3 uM - 6 uM D5, as indicated) and trace 5' $^{32}P$ labeled substrates (0.25 - 1 nM). Initiation of the reaction, measurement of $k_{obs}$ values and single turnover kinetic analysis were conducted as described previously [Michels and Pyle, 1995] with one exception: Preliminary results from fluorescence experiments showed that D1 can slowly fold and bind to the substrate without D5. To ensure that both rS and chimeric substrates reached binding equilibrium with ribozyme before reaction, substrates containing single deoxynucleotides were preincubated for 20 minutes with D1 or D123 prior to the initiation of the cleavage reaction by addition of D5. Appropriate amounts of substrate and D1 were mixed into 4 ul of 80 mM MOPS pH 7.5, and heated at 95°C for 1 minute to denature potentially misfolded structure. After cooling to 42°C, 8 ul concentrated salt in 80 mM MOPS pH 7.5 was added to this D1/S mix to give a final concentration of 1M KCl and 100 mM $MgCl_2$. The solution was then incubated at 42°C for 20 minutes. The reaction was initiated by the addition of 8 ul D5 mixture, preincubated in parallel under the same salt

concentration. This additional procedure was merely a precaution, however, since control experiments showed that for rS and mutant substrates, preincubation of D1 or D123 with substrate results in the same kinetic parameters as analyses without preincubation, performed as described previously [Michels and Pyle, 1995].

Some of the ribozymes described herin are capable of catalytic activity at physiological conditions. Kinetic parameters of a ribozyme which contains domain 1 and domain 3 (D13) is compared with the kinetic parameters of a ribozyme that contains domain 1, domain 2 and domain 3 (D123) (Figures 25C and 25D). Salt concentrations used are the following: 10mM $MgCl_2$, 25mM $MgCl_2$ and 50 mM $MgCl_2$. These low salt conditions are used in the absence of any other monovalent cations such as ammonium or potassium. It is common to include such monovalent cations at molar concentrations in reaction mixtures in order to effect catalytic activity. However, these compounds do not require such salt concentrations. The rate for the D13 ribozyme at 10mM $MgCl_2$ is 9.5 min, at 25 mM $MgCl_2$ is 22.9 min and at 50 mM $MgCl_2$ is 518 min (Fig. 25C). The rate for D123 ribozyme at 25 mM $MgCl_2$ is 12 min and at 10 mM $MgCl_2$ is 54 min. Thus, these compounds are capable of functioning catalytically at physiological conditions.

**Determination of individual kinetic parameters:** In order to determine $K_m$ and $k_{cat}$ for reaction of substrates with the D1 ribozyme, $k_{obs}$ was monitored at varying concentrations of D1, while holding [D5] at saturation. The data fit an equation describing bimolecular association, from which the kinetic parameters $K_m$ and $k_{cat}$ were extracted. The data were quantitated and analyzed as described previously [Michels and Pyle, 1995; and Pyle and Green, 1994]. In order to determine if, in each case, $k_{cat}$ represented $k_{chem}$, it was necessary to conduct parallel pH/rate profiles of the reaction. Values of $k_{cat}$ for rS, -1dC and -2dU substrates

were compared at pH 6.5, 7.0 and 7.5, in 80 mM MOPS, 100 mM $MgCl_2$ and 2 M KCl. MOPS was chosen because it maintains high buffering capacity throughout this pH range. In addition, MOPS maintains a stable pH over a wide temperature range. For these experiments, the KCl concentration was raised to 2 M in order to accurately quantitate the slow rate of -1dC cleavage at pH 6.5. Doubling the KCl concentration speeds up the reaction by approximately 10 fold for all substrates, at all pH values. With the exception of -1dC cleavage at pH 6.5, which is too slow to measure, the same pH/rate behavior is seen for all three substrates (with lower precision) at 1 M KCl.

**Mapping of Cleavage Products and Chemical analysis of 3'-ends:** Partial alkaline degradation and enzymatic digestions with nucleases T1, *B. Cereus* (Pharmacia), and P1 (Boehringer) were carried out as described [Michels and Pyle, 1995]. In order to examine whether cleavage of the -1dC substrate was indeed occurring at the deoxyribose cleavage site, 5' $^{32}P$ labeled substrates, rS and -1dC, were cleaved to 50% completion, under conditions described above. The excess salt was removed with oligo purification cartridges (OPC, from Applied Biosystem, Inc.). The reaction products were then oxidized with 0.1 M $NaIO_4$ in 0.1 M sodium acetate pH 5.0 for 90 minutes at room temperature in the dark [Stephenson and Zamecnik, 1967]. Excess $NaIO_4$ was removed using an OPC cartridge. The oxidized mixture was then reacted with 0.1 M Lucifer Yellow hydrazide (Lucifer Yellow CH, Aldrich) in 0.1 M sodium acetate pH 5.0 at room temperature in dark for approximately four hours. If the 3'-terminus of an oligonucleotide is a ribose, reaction with $NaIO_4$ will oxidize the terminal sugar to a dialdehyde that reacts readily with hydrazine dyes [Stephenson and Zamecnik, 1967; and Hansske et al., 1974], such as Lucifer Yellow. Any 3'-terminal deoxyriboses are resistant to attack of $NaIO_4$. After passing the dye reaction through an OPC cartridge to remove excess Lucifer Yellow, the mixture was

combined with denaturing dye, and loaded onto a 20% denaturing polyacrylamide gel. Attachment of Lucifer Yellow to the 3'-terminus of an oligonucleotide substantially reduces its mobility in a denaturing gel; the identity of 3'-terminal sugars was then analyzed by autoradiography of the products after electrophoresis.

RESULTS:

Preliminary Characterization of Deoxynucleotide Rate Effects:

Effects of block substitution on reaction rate: In order to locate regions of important 2' hydroxyl groups, chimeric substrates were synthesized in which the cleavage site (-1C) was always a ribose and blocks of surrounding residues were substituted with deoxynucleotides. The blocks of deoxynucleotides were clustered into four functional regions: IBS2 (-16 to -11), link (-10 to -8), IBS1 (-7 to -2) and intron (+1 to +7) (Scheme 2). The IBS2 and IBS1 are stretches of nucleotides that pair with complementary sequences on the D1 ribozyme [Michel and Ferat, 1995; and Michel et al., 1989]. The "link" region is a three nucleotide spacer between these paired sequences and the "intron" region is composed of seven nucleotides 3'-to the ribozyme cleavage site (see experimental section). The chimeric substrates were tested using single turnover cleavage reactions under saturating ribozyme conditions (100 nM D1, 3 uM D5), and cleavage rates were compared to those of an all-ribose substrate (rS) (Table 3, column 1).

Table 3. Effects of block deoxyribose substitutions on ribozyme rate.

| | 1[a] | 2 | 3 |
|---|---|---|---|
| Modified Substrate | Effect of dN (rel rate)$^{-1}$ | 250 nM D1[b] (rel rate)$^{-1}$ | 6μM D5[b] (rel rate)$^{-1}$ |
| | | | |
| d(1BS2) | 3.2±0.51 | 1.4±0.22 | 2.9±0.46 |
| D(1BS1) | inactive | inactive | inactive |
| d(link) | 1.4±0.22 | ND | ND |
| d(intron) | 3.3±0.53 | 3.3±0.53 | 3.1±0.50 |

[a]Conditions used: [D1] = 100 nM and [D5] = 3μM, the lower limit for saturation of the unmodified rS substrate. See Table 4 for $K_{obs}$ of rS. Relative rates are $K_{obs}$ for the chimeric substrates divided by that of rS. Variance calculated to 95% confidence, from multiple trials on a single derivative. ND, not determined.

[b]Rate relative to rS cleavage under equivalent conditions. Except for d(IBS1), the block-deoxynucleotide substrates showed small rate decreases relative to rS, indicating that the 2' hydroxyl groups within IBS2, the linker and intronic nucleotide sections of the substrate are not essential for catalytic activity. Within experimental error, the d(link) substrate was as active as rS, so it was excluded from further investigation. The rate of d(IBS2) cleavage was ~ 3 fold slower than rS. That this effect was due to well-documented decreases in hybrid stability [Bevilacqua and Turner, 1991; Martin and Tinoco, 1980; and Hall and McLaughlin, 1991] was supported by restoration of d(IBS2) cleavage rates at higher concentrations of enzyme components (Table 3, column 2). Conversely, the three-fold difference in cleavage rates between rS and d(intron) substrates remained the same at higher concentrations of both D1 and D5 (Table 3, columns 2 and 3), suggesting that the decrease of cleavage rate in d(intron) is due to effects on the chemical rate or some effect other than decrease in binding affinity to either component of the ribozyme.

Under all conditions, d(IBS1) was inactive within the experimental window. Given that this oligonucleotide has a ribose at the cleavage site, there are at least two explanations for its lack of activity: Either important 2'-hydroxyl groups lie within the IBS1 region, or a DNA/RNA hybrid duplex adjacent to the cleavage site is destabilizing for other reasons. The latter possibility is reasonable given that DNA/RNA duplexes are thermodynamically destabilized [Hall and McLaughlin, 1991], their microstructural features can differ slightly from RNA/RNA duplexes [Egli et al., 1993] and they can deviate significantly from A-form geometry and pitch, an extreme case being the recent crystal structure of a B-form DNA/RNA helix [Chen et al., 1995]. But without a signal for activity, it was impossible to differentiate these models and determine which, if any 2' hydroxyl groups of IBS1 were important in ground state or transition-state stabilization.

**Effects of single deoxynucleotide substitutions within IBS1:** To further investigate the importance of 2' hydroxyl groups within IBS1 and at the cleavage site, rS was substituted individually with single deoxynucleotides, and the mono-substituted chimeras were tested using single turnover reactions at saturating ribozyme conditions (Figure 16). The slowest cleavage rate is observed for a substrate containing a deoxyribose at the cleavage site (-1dC). An observed rate of 0.0013 $min^{-1}$ indicates that the 2' hydroxyl group at the cleavage site is the most important of all 2' hydroxyl groups along the substrate. However, this is only a 16 fold decrease relative to the rate of rS cleavage.

Cleavage of the -2dU substrate proceeded with a $k_{obs}$ of 0.0042 $min^{-1}$, which represents a 5-fold decrease in rate relative to rS. Within experimental error, single deoxy substitutions at all other positions within IBS1 (-3 to -7) had only minor effects on rate (Fig. 16). However, as observed for a block deoxynucleotide substitutions of IBS2,

when the -2 to -7 IBS1 nucleotides are simultaneously converted to deoxyribonucleotides, there are likely to be additional effects due to duplex weakening that raise the D1-S $K_m$. Taken together, the single-deoxynucleotide investigations reveal that the most important 2'-hydroxyl groups along the substrate are the 2'-hydroxyl at the cleavage site (-1C) and one that is adjacent to it (-2U) in a region that is normally considered the 5'-exon sequence. Under the single-turnover, saturating ribozyme conditions used in these experiments, the rates were likely to represent effects on the chemical step of the reaction. This was not definitively concluded until further analysis was performed, establishing the rate-limiting step for $k_{obs}$ in each case. Before that could be accomplished, however, it was important to establish that cleavage was taking place at the proper site on the chimeric substrates.

Characterization of Reaction Products:

**Mapping the cleavage site of the -1dC substrate:** One reason that cleavage of a deoxynucleotide linkage might appear to be efficient is that the ribozyme might "skip" the modified linkage and choose an adjacent ribose phosphodiester moiety for attack [Kahle et al., 1993]. As a first approach to investigating that possibility, the cleavage products from the D1/D5 ribozyme reactions were mapped using nucleases and alkaline hydrolysis, followed by high-resolution sequencing (Fig. 17A) [Kuchino and Nishimura, 1989]. Reaction of -1dC with alkali (lane 11) results in a blank space at the position corresponding to mobility of reaction product (a 17-mer, lane 9).

To more precisely evaluate the mobility of the cleavage products, the substrates and products were mapped by partial P1 nuclease digestion, which cleaves both DNA and RNA linkages, resulting in fragments containing the same 3' hydroxyl and 5' phosphate termini observed upon reaction with the ribozyme [Kuchino and Nishimura, 1989]. The

products of ribozyme cleavage of both the rS (lane 6) and the -1dC (lane 9) substrates were found to migrate at the identical position as the 17-mer generated by partial P1 nuclease digestion of each substrate (lanes 7 and 8, respectively). Taken together, the mapping studies suggest that cleavage of the -1dC substrate occurs at the proper sequence and is not the result of attack at an adjacent ribose linkage.

High-resolution mapping of reaction products can be complicated by many factors. It is important to directly probe the terminus of the cleavage product, since correct cleavage of the -1dC substrate will generate a product with a 3' -deoxy terminus. $NaIO_4$ oxidation of an oligonucleotide with a 3' ribose terminus will convert the ribose sugar ring to a dialdehyde, which can be labeled with a hydrazine reagent [Hansske et al., 1974], such as Lucifer Yellow CH hydrazide. The labeled oligonucleotide can be distinguished from the unlabeled one because it migrates more slowly on denaturing gels. An oligonucleotide product with a 3' deoxy terminus would be resistant to oxidation/dye labeling and would not show a mobility shift. In this experiment, oxidation and labeling was done on a 50:50 mixture of substrate and product. The precursor oligonucleotide served as an internal control, since this component always possesses a 3' ribose terminus and should show a mobility shift if the above scheme is correct. The results of this analysis are shown in Figure 17B. For the rS substrate, after oxidation and labeling, both substrate and product shifted up (lanes 3, 4), indicating that both of them have 3' ribose termini. For the -1dC substrate, the substrate mobility shifted after oxidation and labeling, while the product mobility was unchanged (lanes 6,5). This indicates that the cleavage product of -1dC has a 3' deoxynucleotide terminus and that the -1dC substrate is cleaved at the correct site.

Mechanistic Analysis of Deoxynucleotide Rate Effects:

**pH/rate analysis of the chimeric substrates:** The kinetic effects of deoxynucleotide substitution could result from a number of factors: including effects on ribozyme conformation, binding or on the chemical step of catalysis. Furthermore, there have been several reports that deoxynucleotide substitution can change the rate-limiting step of reaction [Smith and Pace, 1993; and Herschlag and Cech, 1990], making comparisons of uncharacterized $k_{obs}$ or $k_{cat}$ values relatively difficult to interpret. Therefore, it was essential to investigate whether $k_{obs}$ for cleavage of rS and -1dC under saturating D1/D5 conditions was governed by the same rate-limiting step. That cleavage of rS is limited by chemistry under these conditions was established previously using a comprehensive kinetic analysis, together with observation of a predicted "thio-effect" on the rate of reaction. As in other hydrolytic cleavages of group II intron substrates, this $k_{chem}$ is expected to vary logarithmically with pH [Pyle and Green, 1994].

Values of $k_{obs}$ for the rS, -1dC and -2dU substrates were obtained at several pH values using the same buffer (Fig. 18). In the pH range tested (pH 6.5 to pH 7.5), both rS and chimeric substrates have a remarkably linear pH/rate profile, in which the rate changes logarithmically with each pH unit. The slope of each pH/rate plot is very close to 1, being 0.87 for rS, 0.94 for -1dC, and 0.85 for -2dU. A linear pH/rate profile with slope of one is regarded as evidence that chemistry is rate-limiting [Herschlag et al., 1993; Pyle and Green, 1994; Dahm et al., 1993; and Khosla and Herschlag, 1994] and that some form of general base catalysis is occuring [Fersht, 1985]. The fact that all three substrates elicit this type of behavior suggests that $k_{chem}$ is limiting in each case and that their cleavage rates can be directly compared.

**Effects of deoxy nucleotide substitution on the binding**

between substrate and ribozyme: The previous pH/rate experiment indicates that 2'-hydroxyl groups at -1C and -2U directly affect the chemical rate of rS cleavage. Although transition-state effects are evident, this does not rule out the possibility that ground state binding might also be affected by the deoxynucleotide substituents, since the large excess in ribozyme concentration will mask, to a certain degree, small changes in binding affinity between chimeric substrates and the ribozyme. To evaluate effects on binding, the rS, -1dC and -2dU substrates were cleaved at varying D1 concentrations. The $k_{cat}$ and $K_M$ values for cleavage of each substrate were calculated from the dependence of $k_{obs}$ on D1 concentration, as described previously [Michels and Pyle, 1995]. The results are shown in Fig. 19 and Table 4. The $k_{cat}$ and $K_m$ values for rS are in close agreement with the previous published results (Fig. 19A) [Michels and Pyle, 1995].

Table 4. Kinetic parameters for D1/D5 ribozyme cleavage of the -1dC and -2dU chimeric oligonucleotide substrates.

| | $k_{cat}$[b] | $K_m$[b] | $k_{cat}/K_M$[c] |
|---|---|---|---|
| Substrate[a] | ($10^{-2}$ $min^{-1}$) | (nM) | ($10^6$ $M^{-1}$ $min^{-1}$) |
| | | | |
| rS(all-RNA) | 2.5±0.070 | 4.9±0.66 | 5.1±0.83 |
| -1dC | 0.19±0.0060 | 14±1.7 | 0.14±0.021 |
| -2dU | 0.51±0.029 | 10±2.4 | 0.51±0.15 |

[a]Only substitutions that had significant effects on reaction rate are shown.

[b]Kinetic parameters and standard errors are derived from fits of the data shown in Figure 19. Kinetic data on the other chimeric substrates is presented in Figures 1,6 and Table 3.

[c]$k_{cat}/K_M$ errors were calculated by propagation of the $K_M$ and $k_{cat}$ errors.

As observed in the previous experiments, the -1dC substrate has a 13-fold decrease on the chemical step ($k_{cat}$), but also has a 3-fold decrease on the ground state binding ($K_m$), resulting in a 36 fold effect on $k_{cat} / K_m$ (Fig. 19B, Table 4). The -2dU substrate has a 5-fold decrease in the chemical step and a 2-fold decrease in the ground state binding (Fig. 19C, Table 4), which may be attributed to duplex destabilization caused by deoxynucleotide substitution. The $k_{cat} / K_m$ for the -2dU substrate is therefore 10-fold lower than that of rS. For both the chimeric substrates, the small decreases in binding affinity ($K_m$) observed here confirm that previous rate comparisons were performed under conditions in which the ribozyme components were indeed saturating. This is also supported by the fact that, within experimental error, the decrease in $k_{cat}$ between chimeric and rS substrates (13 fold for -1dC, 5 fold for -2dU, respectively) agrees with the decrease (16 fold for -1dC, 5 fold for -2dU) in the apparent rates obtained during the previous kinetic comparisons (Fig 16).

Analogous rate effects for a fast group II ribozyme.

One can modify the minimal D1 ribozyme construct by adding back intron domains 2 and 3 in-cis (D123), to generate a ribozyme that cleaves RNA with high efficiency. Under standard reaction conditions in which D123 (100 nM) and D5 (3 $\mu$M) are in excess, this D123 ribozyme cleaves the rS substrate with an apparent $k_{obs}$ of 0.88 $min^{-1}$ (Fig. 20). This is 33-fold faster than the D1 ribozyme under the same conditions, and comparable to the rate of other ribozyme systems. D123 cleaves the -1dC substrate with a $k_{obs}$ of 0.17 $min^{-1}$ (Fig. 20), which is only five-fold slower than cleavage of the ribose linkage and 90-fold faster than deoxynucleotide cleavage by D1. In both cases, cleavage occurs at the proper site (Fig. 17C). Therefore, faster group II intron ribozymes cleave both DNA and RNA linkages with higher efficiency.

There is a smaller experimental window for detecting D1 reactivity, which is radically perturbed upon introduction of a hybrid EBS1/IBS1 duplex immediately adjacent to a ribose cleavage site (as in dIBS1, Table 2). Therefore , it was not possible to detect D1 ribozyme cleavage of an all-DNA oligonucleotide (dS). By contrast, the D123 ribozyme cleaves dS with a second-order rate constant of 2.6 ± 1.4 x $10^3$ $M^{-1}$ $min^{-1}$ under standard reaction conditions using a constant [D5] of 3 μM. Together with effects from single deoxynucleotide substitutions, this result indicates that any barrier to cleavage of entirely DNA molecules is likely to be due to global differences between RNA/RNA and DNA/RNA duplexes and not to a specific role for 2'-hydroxyl groups in the ribozyme mechanism.

The apparent five-fold preference of D123 for ribose over deoxynucleotide linkages compared to the sixteen-fold preference observed with D1 does not actually mean that D123 is less capable of differentiating a ribose from a deoxy linkage. The ratio between the two cleavage rates by D123 is smaller because they have different rate-limiting steps. The cleavage rate of -1dC by D123 (0.17 $min^{-1}$) is limited by the rate of chemistry, as demonstrated through pH/rate and other mechanistic analyses. However, the rate of rS cleavage by D123 (0.88 $min^{-1}$) is not limited by the chemical rate of catalysis. Unlike all of the other processes reported in this paper, this rate is not pH dependent and it has a similar magnitude to rates of folding processes observed in stopped-flow fluorescence experiments performed under identical reaction conditions. Preliminary estimates for the chemical rate of rS cleavage by D123 are approximately 3 $min^{-1}$, which would indicate that D123 has a similar ~15-fold preference for ribose observed for the D1 ribozyme. Taken together, these results suggest that significant transition-state stabilization can be achieved by fast group II intron ribozymes without engaging 2'-hydroxyl groups for high levels of transition-state

stabilization. It is therefore clear that ribozymes in mechanistic class B do not necessarily exploit the stabilizing effects of the 2'-hydroxyl group at the cleavage site.

**Discussion**

At this time, we have only a modest understanding of the means by which RNA behaves as an enzyme or potentiates chemical events. There are very few natural ribozymes in existence, and only three in the mechanistic class B to which the group II intron belongs [Pyle, 1993; and Cech, 1993]. For this reason, each ribozyme is an important resource, providing a distinctive window on mechanisms for transition-state stabilization by RNA. Group II intron active sites are particularly fascinating because they catalyze a variety of sophisticated reactions, including forward branching [Peebles et al., 1986; Schmelzer and Schweyen, 1986; and van der Veen et al., 1986], reverse branching [Mörl and Schmelzer, 1990; Augustin et al., 1990; and Chin and Pyle, 1995], site-specific hydrolysis [Michels and Pyle, 1995; Jarrell et al., 1988A; Jarrell et al., 1988B; Pyle and Green, 1994; and Podar et al., 1995], polymerization [Müller et al., 1993], editing [Müller et al., 1993] and other reactions on RNA and DNA [Mörl et al. 1992]. To understand more about the basis for catalysis by group II intron active-sites, we constructed a series of group II intron ribozymes that cleave RNA and DNA oligonucleotides in a manner that reflects factors essential for RNA folding and transition-state stabilization. In the most minimal system, a group II intron ribozyme was constructed of three pieces in which D1 behaves as the ribozyme, D5 as a catalytic cofactor, and a short oligonucleotide complementary to the EBS sequences acts as the substrate (Scheme 2). This ribozyme catalyzes a single-step reaction, cleaving the oligonucleotide substrate into two pieces using water or activated hydroxide as the nucleophile. Previous studies have delineated a kinetic

framework for activity by this ribozyme and identified a rate-limiting chemical step of reaction [Michels and Pyle, 1995]. This accessibility of its chemical rate constant renders this ribozyme useful for studies of transition-state stabilization. Because the D1/D5 ribozyme cleaves short oligonucleotides, these substrates can be generated on an automated DNA/RNA synthesizer, making it possible to incorporate unnatural nucleotides and single-atom changes for probing the chemical mechanism. There are two major results from these structure/function studies: First of all, the data indicate that a 2'-hydroxyl group at the cleavage site has only a small role in lowering the activation barrier for reaction. This effect is observed also for fast variants of the group II intron ribozyme. Second, the 2'-hydroxyl groups along the substrate backbone do not form an extensive network of sequence-independent backbone tertiary interactions upon binding to the ribozyme.

**The chemical role of 2' hydroxyl groups on the substrate:** Previous studies have examined the role of ribose sugars at the site of cleavage by other ribozymes in mechanistic class B [Herschlag et al., 1993; Smith and Pace, 1993; Herschlag et al., 1993; Perreault and Altman, 1992; Robertson and Joyce, 1990; Herschlag and Cech, 1990A; Herschlag and Cech, 1990B; and Kleineidam et al., 1993]. These studies have shown that a 2'-hydroxyl group adjacent to the scissile bond is important in promoting catalysis. Single deoxynucleotide substitution at the cleavage site of the *Tetrahymena* ribozyme (derived from a group I intron) resulted in a 590-fold decrease in the chemical rate of reaction [Herschlag et al., 1993A; and Herschlag et al., 1993B]. Analagous studies on the RNA component of RNase P resulted in a relative chemical rate difference of 3400 fold [Smith and Pace, 1993]. It was therefore surprising that, in a family of group II intron ribozymes, the absence of a 2' hydroxyl group at the cleavage site caused only a 16 fold reduction in chemical rate (Fig. 21). This 16-fold differential

between ribose and deoxynucleotide substitution is consistent with the magnitude of an "induced effect" [Herschlag et al., 1993A; Herschlag et al., 1993B; and Exner, 1978], in which the electron withdrawing property of the 2′ hydroxyl group plays a role in stabilizing the adjacent 3′ oxyanion leaving group that develops in the transition state. This effect stems from the fact that the -OH is a better electron withdrawing group than the -H. In a reaction in which negative charge develops on a nearby atom (3′-oxygen), the -OH group will better stabilize the transition state than the -H group, thus speeding up the reaction.

Although unexpected, there is no fundamental principle against a limited role of a 2′ hydroxyl group at the cleavage site. Since the 2′ hydroxyl group does not participate as nucleophile or leaving group, the chemical mechanism of ribozymes in mechanistic class B does not implicitly require its presence (Scheme 1) [Pyle, 1993]. The closest example of this type of behavior was observed in studies of pre-mRNA splicing, a fact that is significant because group II introns are widely viewed as possible evolutionary predecessors of the eukaryotic spliceosome. In that work, it was reported that deoxynucleotide substitution at the 5′ splice site had no effect on either step of splicing [Moore and Sharp, 1992]. While the chemical step of catalysis is unlikely to be rate-limiting in that complex system, it may be significant that there was no apparent block to reactivity, especially considering that base mutations at that site have strong effects on the apparent efficiency of splicing [Newman et al., 1985; and Parker and Siliciano, 1993]. There is also precedent for group II intron-catalyzed attack on DNA linkages. Although it was reported to be very inefficient, the free bI1 intron (a different group II intron from yeast mitochondria) was observed to promote cleavage of a DNA substrate [Mörl et al., 1992].

Small rate effects were observed for the 2' hydroxyl group located one nucleotide 5'-from the cleavage site, at position -2U. This 2'-hydroxyl group does not appear to participate in specific ground state contacts. However, deoxynucleotide substitution at this position leads to a 5-fold reduction in $k_{chem}$. This is interesting because it implies that 2' hydroxyl groups at nucleotides adjacent to the cleavage site can be directly involved in the transition state. Perhaps the 2' hydroxyl group at position -2U is involved in a tertiary contact that generates tension, or strain on the adjacent scissile phosphodiester linkage. This 2'-hydroxyl group may be close enough to act as an outer-sphere ligand for a catalytic metal ion.

**Rapid deoxynucleotide cleavage by a group II intron ribozyme:**

D1 cleavage of an all-ribose substrate is slower than that of many known ribozymes. The degree to which the D1 ribozyme is considered "slow" must be put in relative terms since it still provides a $10^7$-fold rate-enhancement over the uncatalyzed cleavage reaction [Michels and Pyle, 1995]. However, compared to ribozymes derived from group I introns and RNase P, its chemical rate for oligonucleotide cleavage is inefficient [Herschlag and Cech, 1990; and Beebe and Fierke, 1994]. One explanation for this inefficiency is that the D1/D5 ribozyme does not contain all of the active-site functionalities necessary for efficient cleavage, including contacts to the 2'-hydroxyl group at the cleavage site. This effect might quantitatively account for the ~5 kcal/mol by which this ribozyme is chemically slower than others [Herschlag et al., 1993A; Smith and Pace, 1993; Herschlag and Cech, 1990; and Beebe and Fierke, 1994]. It was therefore important to determine if the 2' hydroxyl group at the cleavage site contributes to catalysis by a fast group II intron ribozyme.

The D1/D5 ribozyme can be altered such that it catalyzes

highly efficient oligonucleotide cleavage. In particular, when Domain 3 (~80 nucleotides in ai5g) is covalently attached to Domain 1, one observes an apparent cleavage rate approaching 1 $min^{-1}$ for rS. For the fast variants, such as D123 described here, deoxynucleotide cleavage is also efficient, proceeding with a rate of 0.17 $min^{-1}$ (Fig. 20). This is 90-fold faster than deoxynucleotide cleavage by the D1/D5 construct. The D123 ribozyme can also cleave an all-DNA oligonucleotide, apparently overcoming difficulties associated with a totally hybrid IBS/EBS duplex. The ease of DNA cleavage and the similar rates for rS and -1dC cleavage by D123 suggest that this ribozyme, like its slower relative D1, does not utilize the 2'-hydroxyl at the cleavage site for extensive transition-state stabilization.

The $k_{cat}$ value of 0.17 $min^{-1}$ for -1dC cleavage by D123 is orders of magnitude faster than deoxyribonucleotide cleavage by any known natural ribozyme [Lesser et al., 1990; and Walsh, 1979]. It is similar to the $k_{cat}$ for cleavage of DNA by EcoRI (0.72 $min^{-1}$) and only an order of magnitude slower than the individual rate constant for the bond-breaking step [Kunkel et al., 1991; and Scaringe et al., 1990]. The $k_{cat}$ for -1dC cleavage by D123 is only two-fold slower than the rate observed for the best mutant group I intron ribozyme that was evolved through 27 generations of mutation and selection to cleave a DNA linkage [Tsang and Joyce, 1994]. However, the $k_{cat}/K_m$ value for the D123 ribozyme is $\geq$ 3.2 x $10^7$ $M^{-1}min^{-1}$ (estimating a $K_m$ of 5 nM for S binding to D123), a value which is 10-fold faster than $k_{cat}/K_m$ of the best DNA-cleaving ribozyme derived through directed molecular evolution [Tsang and Joyce, 1994]. This result is primarily significant because it was obtained with components of a group II intron active site that have not been selectively optimized in-vitro for increased catalytic activity or for activity specifically directed against a DNA linkage. Although detailed kinetic characterization of D123 and D13 ribozymes is currently under investigation, the results

presented here suggest that oligonucleotide-cleaving ribozymes of the D1 or D123 class, together with catalytic D5, exploit a catalytic strategy in which the 2'-hydroxyl group at the cleavage site is not central to stimulation of catalysis.

Efficient deoxyribose cleavage by a group II intron ribozyme may have implications for the apparent mobility, or migration of group II introns into new genomic positions [Lambowitz, 1989; and Lambowitz and Belfort, 1993]. It is widely speculated that free group II introns can bind and then reverse-splice into new RNA molecules, provided that these share sequence similarity to the parent site [Van Dyck et al., 1995; and Schmidt et al., 1994]. It may also be important to consider that group II introns could attack and integrate into DNA sites, thus becoming directly incorporated into a new genome [Mörl et al., 1992]. Despite the temptation to speculate on a biological role for DNA cleavage by these ribozymes, it is important to note that the catalytic behavior of group II intron ribozymes that cleave oligonucleotide substrates is often different than that observed for cis-splicing and reactions in which the cleavage site is covalently attached to D1 [Podar et al., 1995]. Although they are more difficult to analyze kinetically, cis-cleaving constructs will need to be substituted with deoxynucleotides before strong statements can be made about the function of 2'-hydroxyl groups in specific cis-splicing or reverse-splicing reactions.

**Ground-state effects of 2'-hydroxyls on the substrate:** For the D1/D5 ribozyme, the effect of deoxynucleotide substitutions on ground state binding was measured by changes in the $K_m$ value. This $K_m$ is considered a reasonable measure of D1-S association energy because it was previously shown to equal $K_d$ for binding of rS [Michels and Pyle, 1995]. Because $k_{cat} = k_{chem}$ for each substrate investigated here, it is likely that $K_m=K_d$ for the chimeric substrates as well. $K_m$

values for the -1dC and -2dU substrates are only 3 fold and 2-fold larger than for rS (Table 4). These effects are similar to the magnitude of duplex weakening effect observed previously upon single deoxynucleotide incorporation into an RNA/RNA helix [Bevilacqua and Turner, 1991; and Hall and McLaughlin, 1991], suggesting that these 2' hydroxyl groups do not form tertiary contacts in the ground state. It is also noteworthy that in the group I *Tetrahymena* and RNase P ribozymes, small ground-state effects of ~3 fold were also observed for substitution at the -1 position [Herschlag et al., 1993A; Smith and Pace, 1993; and Herschlag et al., 1993B].

Single deoxynucleotide substitutions within the IBS1 region at positions -3 to -7 resulted in no observable effect on binding or chemistry (Fig. 16, Fig. 21). Being far away from the splice site, it is not surprising that these 2' hydroxyl groups contribute little, if any, energetic stabilization in the transition state. The lack of any major effects on the rate at 100 nM D1 and 3 μM D5, which represent the lower limit for saturation of the rS substrate, also suggests that the ground state has not been substantially disturbed and that there are no ground-state tertiary contacts between any of these 2' hydroxyl groups and functionalities on the ribozyme. A lack of backbone tertiary interactions in oligonucleotide binding by group II intron ribozymes is supported by the fact that ribozyme-substrate binding energies parallel the amount of duplex stabilization one can calculate for base-pairing of the two EBS-IBS interactions [Michels and Pyle, 1995; and Mörl et al., 1992]. This contrasts with the behavior of group I intron ribozymes, in which 2' hydroxyl groups were found to make specific tertiary contacts that provide extra binding energy that stabilizes both the ground and transition states [Pyle et al., 1992; Herschlag et al., 1993; Strobel and Cech, 1993; Pyle and Cech, 1991; and Bevilacqua and Turner, 1991].

**Implications for target sequence-specificity:**

Because 2' hydroxyl groups appear to make minimal contributions to stabilizing S-D1 association, group II introns may be able to obtain a higher level of specificity in substrate recognition. Ribozymes that rely on extensive tertiary interactions with the backbone of their substrates can suffer from a loss of sequence specificity. This is because, regardless of base identity, every residue possesses 2' hydroxyl and phosphate groups. If these functionalities are important in substrate binding, they may undermine the energetic differential between matched and mismatched base-pairings, allowing a mismatched substrate to be "tolerated" long enough in the binding site to undergo cleavage [Herschlag, 1991]. By excluding 2' hydroxyl contacts from the total ribozyme-substrate stabilization energy group II introns may experience an enhanced "signal to noise", in which nonspecific background energy is lowered and the contrast between canonical and non-canonical base pairing stands out.

An exceptionally high level of specificity in substrate recognition by group II intron ribozymes may also result from the presence of two separate substrate binding domains: the EBS1 and EBS2 regions, which bind to noncontiguous portions of Domain 1. By breaking the substrate pairing sequences into pieces and putting linkers between them, it may be possible to target a highly distinctive sequence (in this case, 16 nucleotides including the linker region) without adding so much binding energy that mismatches within the stretch will be tolerated. It is worth noting that targeting a unique site within the human genome requires base-pairing recognition of a site that is at least 16 nucleotides in length [Sambrook et al., 1989]. The actual sequence of the EBS-IBS interactions is apparently unimportant, since the EBS-IBS interactions covary to retain base-pairing in group II introns of different sequence

[Michel et al., 1989]. Therefore, by mutating the EBS sequences, D1 can be made to pair with any IBS sequence and cleave at the appropriate site. Given the results provided here and in other work, it may even be possible to target DNA sequences. Thus, group II intron ribozymes may become an important molecular biological tool for the in-vitro manipulation of DNA and RNA. Furthermore, the enzymatic characteristics described in this work imply that group II intron ribozymes and trans-splicing variants are promising agents for gene therapy at the DNA or RNA level and for the site-directed gene repair of defective mRNA.

Significance

Although group I introns have been the subject of detailed analyses, there have been few mechanistic studies of catalytic activity by group II intron ribozymes. In spite of their biochemical importance, the structural complexity and size of group II introns make them difficult subjects for chemical and enzymological analysis. A kinetic framework for catalysis by several group II intron ribozymes was recently established, providing a basis for detailed studies on the role of specific atoms in the chemical mechanism. Using this framework, we examined the role of 2'-hydroxyl groups on substrates for a family of group II ribozymes. The 2'-hydroxyl group of ribose is central to substrate association and/or cleavage by other ribozymes. By contrast, we observe that substrate 2'-hydroxyls make minor contributions to both transition- and ground-state stabilization by group II intron ribozymes. In fact, cleavage of DNA linkages (which lack 2'-hydroxyl groups) is remarkably efficient for group II intron ribozymes. The results have important implications for ribozyme chemical reactivity and target specificity.

Example 4: Two competing pathways for self-splicing by group II introns: a quantitative analysis of in-vitro reaction rates and product

SUMMARY

Self-splicing group II introns are found in bacteria and in the organellar genes of plants, fungi, and yeast. The mechanism for the first step of splicing is generally believed to involve attack of a specific intronic 2'-hydroxyl group on a phosphodiester linkage at the 5'-splice site, resulting in the formation of a lariat intron species. In this example, we present kinetic and enzymatic evidence that in-vitro there are two distinct pathways for group II intron self-splicing: one involves 2'-OH attack (branching, Fig. 22A) and another involves attack of water or hydroxide (Fig. 22B). These two pathways occur in parallel under all reaction conditions, although either can dominate in the presence of particular salts or protein cofactors. Both pathways are followed by a successful second step of splicing, and either pathway can be highly efficient. We find that the hydrolytic pathway prevails under physiological ionic conditions, while branching predominates at molar concentrations of ammonium ion. The intron is observed to adopt two major active conformations. In order to quantitate their individual reaction rates, we applied a mechanistic model describing biphasic parallel kinetic behavior. Kinetic analysis throughout the investigation reveals that there is no coupling between the unproductive "spliced-exon-reopening" reaction (SER) and hydrolysis during the first step of splicing. Conditions that stimulate branching can promote the SER reaction just as efficiently as conditions that stimulate the hydrolytic pathway. Although there is little evidence that it exists in-vivo, a hydrolytic splicing pathway for group II introns has important implications for the translation of intron-encoded proteins and the inhibition of intron migration into new genomic positions.

INTRODUCTION:

Group II introns are self-splicing RNAs that are distinct in

structure and mechanism from another class of self-splicing RNAs: the group I introns (Cech, 1993). The secondary structure of group II introns can be organized into a specific set of domains (Davies et al., 1982; Michel et al., 1982; Schmelzer et al., 1982), although unlike many other catalytic RNAs, there are few conserved nucleotides or phylogenetic covariations between the functional domains (Michel et al., 1989; Costa & Michel, 1995). Group II introns are found in plants, fungi and yeast (Michel et al., 1989), where many of them are essential for the expression of genes involved in metabolism. In addition, these introns were recently identified in bacteria (Ferat & Michel, 1993; Ferat et al., 1994). Although many group II introns do not contain coding regions, certain of them contain open reading frames for the translation of a unique class of enzymes known as the "maturases" (Lambowitz & Perlman, 1990; Lambowitz & Belfort, 1993). Containing RNase H and reverse-transcriptase-like domains, maturase proteins have been found to be important in intron mobility and have been postulated to stimulate efficient group II intron self-splicing in-vivo. The mechanism of self-splicing and related reactions by active site(s) within group II introns has been the subject of several recent investigations (Padgett et al., 1994; Pyle & Green, 1994; Chin & Pyle, 1995; Michels & Pyle, 1995; Podar et al., 1995a; Podar et al., 1995b). The chemical versatility and structural complexity of group II introns make them interesting subjects for studies of RNA folding and catalysis. In addition, mechanistic parallels with reactions by the eukaryotic spliceosome have led many to speculate that group II introns are evolutionary predecessors of, and therefore simple chemical models of, spliceosomal active sites (Sharp, 1985; Cech, 1986; Jacquier, 1990). Like the investigation reported herein, many of the studies on group II intron mechanism have utilized the ai5g group II intron, which is the fifth intron in the gene for the third subunit of cytochrome oxidase of yeast mitochondria (Peebles et al.,

1986).

The hallmark of self-splicing by group II introns is a mechanistic pathway involving 5'-splice site attack by a bulged adenosine residue within the intron (Fig. 22A). As in nuclear pre-mRNA splicing (Padgett et al., 1984), the 2'-hydroxyl group of this "branch-point" adenosine is the nucleophile during the first step of splicing, giving rise to lariat intron as the product of a complete splicing reaction (Peebles et al., 1986; Schmelzer & Schweyen, 1986; van der Veen et al., 1986). Apparent mechanistic parallels with the spliceosome have significantly influenced the way we think about group II intron reactivity and structure. However, because of their innate importance in the splicing of essential genes in many organisms, and because of potential similarities to RNAs other than those contained in the spliceosome, this study attempts to independently evaluate the mechanism for the first step of self-splicing by group II introns.

In this example, we show that there are two routes for competent self-splicing by group II introns: in one mechanism, the first step of splicing proceeds through branch-point attack (Fig. 22A), while in the other, the first step of splicing is mediated by attack of water or hydroxide ion on the 5'-splice site (Fig. 22B). Previous studies reported the prevalence of the latter pathway under a particular set of reaction conditions (Peebles et al., 1987; Jarrell et al., 1988a; Jarrell et al., 1988b). In this study, we show that competing hydrolytic and branching pathways occur in parallel under all reaction conditions tested, leading us to speculate that there may be an important biological role for the hydrolytic pathway. The data presented here represents the first side-by-side analysis of all the in-vitro ionic conditions known to promote self-splicing, together with several new reaction conditions. Detailed timecourses of self-splicing were

analyzed by applying classical models of parallel kinetics (Fersht, 1985; Jencks, 1987) and accounting for two different conformations of the intron. This approach provides the groundwork for rigorous interpretation of rate data from future structure/function studies on group II introns.

**RESULTS:**

**Qualitative analysis of chemically distinct self-splicing pathways:**

In several buffer salts, we conducted reaction time-courses and inspected the products after separation by PAGE (Fig. 23A). It was previously reported that the predominant splicing products observed in the presence of ammonium salts were lariat intron and spliced exons (Peebles et al., 1987; van der Veen et al., 1987; Jarrell et al., 1988b; Bachl & Schmelzer, 1990; KŸck et al., 1990; Jacquier & Jacquesson-Breuleux, 1991). These RNAs were observed together with a linear intronic species that is frequently labeled as "broken lariat". We also observe splicing in $(NH_4)_2SO_4$ to be the most efficient condition for lariat formation and splicing of exons (Peebles et al., 1987) (band A, Fig. 23A). There is actually a rapid burst in the production of lariat and spliced exons early in reaction, which is more evident from the quantitation below. However, very early in reaction, we also observed evolution of the "broken lariat" species, which increased steadily throughout the course of reaction (band C, Fig. 23A). At 900 minutes of reaction (lane 15, Fig. 23A), the predominant species is still the lariat molecule. Given that the majority of lariat molecules were stable enough to avoid random nicking after this length of time, it is curious that "broken lariat" molecules are evident after only four minutes of reaction. If the rate of conversion of lariat to broken lariat were that fast, it would seem that after 900 minutes, most of the lariat would be broken. This simple observation leads to the following hypothesis: Perhaps the "broken-lariat" band

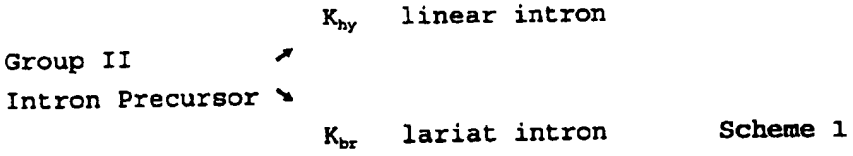
is not the result of lariat degradation. Instead, it is a linear species resulting from hydrolytic, rather than branch-point attack at the 5'-splice site. Given that the 5'-exon product will have the same 3'-hydroxyl terminus, regardless of whether it was produced through branch-point or hydrolytic attack, it should be possible to follow a hydrolytic first step with a second step of splicing. In the end, this reaction would lead to spliced exons and the production of linear intron, much like that observed for group I intron splicing (Cech, 1993).

A hydrolytic pathway, occuring during the first step, was previously proposed to account for the abundant linear intron produced at high concentrations of KCl (Peebles et al., 1987; Jarrell et al., 1988b). We reasoned that, if the pathway could occur in the presence of KCl (which also produces significant amounts of lariat intron), there might be some propensity for this pathway to occur under all conditions. Specific ions would be expected to increase the relative rate of one of the two possible splicing pathways, resulting in apparent dominance of a particular reaction product. To visualize this effect qualitatively, we compared the relative proportions of linear and lariat intron, produced under KCl and $NH_4Cl$ conditions (Fig. 23B and 23C). Like $(NH_4)_2SO_4$, splicing in the presence of $NH_4Cl$ results predominantly in lariat evolution (Band A, Fig. 23C). However, large amounts of linear intron are produced as well (Band C, Fig. 23C). In KCl, which was previously reported to promote formation of linear intron (Peebles et al., 1987; Jarrell et al., 1988b), we observe the production of both linear and lariat species (Fig. 23B). The amount by which lariat dominates over linear products in $NH_4Cl$ is reversed under KCl conditions, and neither product dominates by more than several-fold. This reasoning suggested that the same set of reactions were going on under all of the different reaction conditions. In each case, two pathways for splicing were evident: branch-point attack on the 5'-

splice site competes with hydrolytic attack on the 5'-splice site and, in each case, the free 5'-exon can go on to participate in the second step of splicing.

**Kinetic analysis of two competing splicing pathways:**

If linear and lariat molecules arise from competing mechanisms of cleavage at the 5'-splice site, then self-splicing will follow "parallel-first order reaction kinetics" as defined previously (Fersht, 1985; Jencks, 1987). A first-order treatment is appropriate, in spite of the two-step nature of self-splicing, because intermediate products (such as lariat 3'-exon) do not accumulate in abundance and the first step of self-splicing is rate-limiting (Jacquier and Jacquesson-Breuleux, 1991). Parallel kinetics of self-splicing are illustrated in in Scheme 1:




Scheme 1

Using this form of kinetic analysis, we sought to define reaction efficiency as the actual rate for individual pathways of splicing, i.e. branching and hydrolysis ($k_{br}$ and $k_{hy}$) under a variety of reaction conditions. In order to solve for $k_{br}$ and $k_{hy}$ , one must note three features of parallel first-order reactions (Jencks, 1987) (1) All products will evolve at the same total rate ($k_T$), which is equal to the sum of the individual rates ($k_{hy}$ and $k_{br}$). (2) The ratio of the different products ($f_{hy}$ / $f_{br}$, defined in Materials and Methods) will be a constant through time. (3) If parallel kinetics apply in this case, then the sum of $f_{hy}$ and $f_{br}$ will equal 1 even if equations 1-3 (Materials and Methods) are fit independently (as performed here). Given these characteristics of a parallel reaction, it should not be possible to obtain $k_{hy}$ or $k_{br}$ directly by monitoring evolution of a single reaction product. Instead, one must

solve for $k_{hy}$ and $k_{br}$ by determining the relative fractional amounts of each product and using these values to correct $k_T$ (equations 4 and 5).

**Kinetics of self-splicing in 100 mM $MgCl_2$:** The first data sets to be analyzed were those reaction conditions that included $Mg^{2+}$, but lacked monovalent ion. Commonly referred to as "low-salt" conditions, these are known to stimulate a slow self-splicing reaction (Peebles et al., 1986; Peebles et al., 1987). The most common "low salt" condition currently in use is 100 mM $MgCl_2$ (Chanfreau & Jacquier, 1994; Boulanger et al., 1995). It has been reported that mutational sensitivities of splicing under this condition show the strongest parallels to effects observed in-vivo (Boulanger et al., 1995). Extended timecourses under this condition revealed a slow reaction following simple kinetic behavior that best fit equations 1-3.

Important features of the time-course include the following: First of all, the overall $k_T$ values determined in three separate ways - reaction of precursor, evolution of lariat or evolution of linear intron - are all in close correpondence, differing by less than a factor of two (legend, Fig. 24A). This correspondence implies that the evolution of lariat and linear intron follow parallel first-order kinetics and thereby evolve simultaneously and independently from precursor. The evolution of spliced and free exons also resulted in a rate equivalent to $k_T$ (data not shown). However, because the exons were smaller than intronic products, they were less radioactive and thus less accurate for tracking the course of the splicing reaction. In all cases, the evolution of intronic products has been correllated with the same extent of exon splicing.

As in other reactions that did not contain molar quantities of ammonium salts, evolution of linear intron dominated over lariat intron (Fig. 24A). The latter observation, together

with poor fit to the sequential reaction model (Tinoco et al., 1978), makes it unlikely that a sequential pathway for linear formation from lariat is the operative mechanism. As will be seen by comparing these data with reaction conditions involving high amounts of monovalent cation (below), the entire population of molecules in $Mg^{2+}$ salts alone reacts slowly, with rates similar to a subset of molecules that react slowly in the presence of monovalent ions (Table 5). For this reason, all of the rate data for the homogeneous population of molecules reacting in 100 mM $Mg^{2+}$ or 10 mM $Mg^{2+}$, is located in section B of Table 1. The overall rate of splicing observed under these conditions was 0.0049 $min^{-1}$. From this value, and from the asymptotic coefficients for linear and lariat accumulation ($f_{hy}$ and $f_{br}$, Table 5), we calculated $k_{hy}$= 0.0040 $min^{-1}$ and $k_{br}$ = 0.0009 $min^{-1}$ (Table 5, legend to Fig.24A).

Table 5 Rate Constants[a] for Self Splicing

| Ionic Rxn | Cndtn[c] | 100mM $MgCl_2$ | 10mM $MgCl_2$ 5nM protamine | 500mM KCl 100mM $MgCl_2$ | 500 mM NH4Cl 100mM $MgCl_2$ | 500 mM $(NH_4)_2SO_4$ 100mM $MgCl_2$ |
|---|---|---|---|---|---|---|
| $k^A_T$ | ($min^{-1}$)[d] | - | - | 0.028 | 0.038 | 0.22[e]/ 0.047[f] |
| $k^A_{br}$ | ($min^{-1}$) | - | - | 0.0063 | 0.038 | 0.22[e]/ 0.047[f] |
| $f^A_{br}$ | ($min^{-1}$) | - | - | 0.093 | 0.38 | 0.13[g]/ 0.26[g] |
| $k^A_{hy}$ | ($min^{-1}$) | - | - | 0.022 | 0 | 0 |
| $f^A_{hy}$ | ($min^{-1}$) | - | - | 0.32 | 0 | 0 |
| $k^B_T$ | ($min^{-1}$) | 0.0049[b] | 0.0029[b] | 0.0042 | 0.0039 | 0.0028 |
| $k^B_{br}$ | ($min^{-1}$) | 0.0009 | 0.0004 | 0.0011 | 0.0006 | 0.001 |
| $f^B_{br}$ | ($min^{-1}$) | 0.18 | 0.14 | 0.16 | 0.10 | 0.22 |
| $k^B_{hy}$ | ($min^{-1}$) | 0.0040 | 0.0025 | 0.0032 | 0.0032 | 0.0018 |
| $f^B_{hy}$ | ($min^{-1}$) | 0.75 | 0.77 | 0.48 | 0.51 | 0.39 |
| $\Sigma f^i$ | (i=A, B) | 0.93 | 0.91 | 1.05 | 0.99 | 1.00 |

a. The rates reported here should be viewed as observed rates, and not the rate of a specific chemical or conformational step. Although there is evidence to suggest that the hydrolytic rate constants shown here represent a chemical rate and that the burst rate for branching is a lower limit for the chemical step in that reaction, it was not within the scope of this study to dissect specific rate constants along the splicing pathway.

b. Fast and slow populations of molecules refer to the two conformational states identified during kinetic and chromatographic analysis of the precursor RNA. All rate constants for reaction in the absence of monovalent ion (the first two rows) are put in the category of Population B because their rates matched those of Pop. B under all other reactions conditions. However, given the slow rates and limited reaction extent in 10 mM $MgCl_2$ and 100 mM $MgCl_2$, one cannot rule out the existence of a second, less active, conformation under these conditions.

c. In addition to the salts listed, each reaction contained 40 nM MOPS, pH 7.5.

d. The variables $k_T$, $k_{br}$, $f_{br}$, $k_{hy}$ and $f_{hy}$ for a given population (A or B) are the total rate of splicing, the rate of branching, the fraction of molecules that splice through branching, the rate of hydrolytic splicing, and the fraction of molecules that splice through hydrolysis, respectively.

e. The rate of reaction for the 13% of the total population that undergoes a burst of reaction.

f. The remainder of population A that does not burst, but still reacts through branching at a fast rate.

g. The top value (13%, determined independently, Figure 5) represents the fraction of molecules that proceed through a burst. The bottom value represents the remainder of the fast population.

**Kinetics of self-splicing in 10 mM $MgCl_2$:** The first set of ionic conditions ever used to detect group II intron self-splicing were 10 mM $Mg^{2+}$ with 2 mM spermidine (Peebles et al., 1986). Under these conditions, lariat and linear products form in approximately equimolar ratios. However, in our hands, RNA degradation was so severe that it was difficult to perceive the accumulation of either spliced or free exons (lanes 15-17, Fig. 28). The data was of sufficient quality to quantitate an overall rate of

precursor reaction (0.003 $min^{-1}$, data not shown), but degradation made it difficult to accurately partition this rate between relative contributions from linear and lariat pathways. Because of this problem, we found an alternative polyamine for analyzing activity at 10 mM $Mg^{2+}$. We observed that stoichiometric amounts of protamine, a small arginine-rich polypeptide, promoted significant amounts of reaction at 10 mM $Mg^{2+}$, without concomitant RNA degradation. This could be attributable to the formation of a tightly folded RNA structure in the presence of protamine, or an inability to remove traces of nuclease from the spermidine. In the presence of 5 nM protamine, the accumulation of lariat RNA is similar to that observed using 10 mM $Mg^{2+}$ with 2 mM spermidine, although all the linear products of self-splicing (such as linear intron, spliced exons and free exons), which tend to be more susceptible to degradation, can be quantitated because they are readily detected above the noise of random RNA degradation. As in the case of 100 mM $Mg^{2+}$, the reaction obeys parallel kinetics in which the linear intron product dominates over lariat intron product (Fig. 24B). That both pathways lead almost exclusively to the production of ligated exons (rather than free exons) is apparent from product analysis (lanes 18-20, Fig. 27). The overall rate of splicing observed under these conditions was 0.0029 $min^{-1}$ (as seen for 10 mM Mg2+ and 2 mM spermidine). From this value, and from the asymptotic coefficients for linear and lariat accumulation ($f_{hy}$ and $f_{br}$), we calculated $k_{hy}$= 0.0025 $min^{-1}$ and $k_{br}$ = 0.00044 $min^{-1}$. These rates are both approximately two-fold slower than in the presence of 100 mM $Mg^{2+}$.

**Kinetic analysis of self-splicing with high concentrations of monovalent ion:**

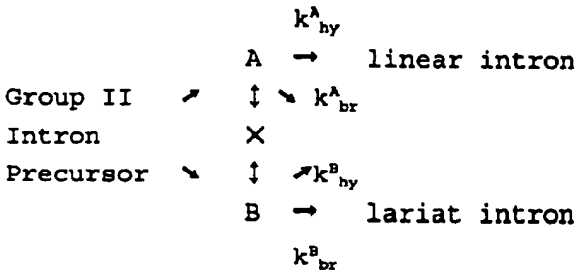
Previous studies reported that monovalent cations had an inhibitory effect on the rate of self-splicing in the presence of 10 mM $Mg^{2+}$ (Peebles et al., 1986), although high concentrations of certain salts had a marked stimulatory

effect on rate in the presence of 100 mM $Mg^{2+}$ (Peebles et al., 1987). Presumably this is because monovalent ions facilitate folding at high concentration. However, at high concentration, monovalent ions may be able to compete with low concentrations of catalytically essential divalent ions in the active site. Thus, one might be required to correspondingly increase the concentration of divalent ion in order to exploit the stimulatory effects of monovalent ions. Several studies defined three salts that are particularly stimulatory for splicing in the presence of 100 mM $MgCl_2$ at concentrations from 0.2M to 2M: KCl, $NH_4Cl$ and $(NH_4)_2SO_4$ (Peebles et al., 1987; Schmelzer & Muller, 1987; van der Veen et al., 1987; Jarrell et al., 1988b). While these are the major salts employed in group II self-splicing studies today, we also examined several others, including potassium acetate.

**Biphasic kinetic behavior:** Although they appear to favor different pathways during the first step of splicing (Fig. 24), plots of the overall self-splicing reaction in $(NH_4)_2SO_4$, KCl and $NH_4Cl$ are remarkably similar. Each of them shows pronounced biphasic behavior (Fig. 25) with a transition at approximately 50% reaction extent. That is, the data for each timecourse fall on two straight lines, which intersect at ~ 50% reaction completion. Because this apparent switch from fast to slow reaction occurs at approximately the same reaction extent under such radically different reaction conditions, it suggests that there are two major conformations of the precursor RNA, resulting in two populations of molecules: a fast population and a slow population. The slope of the line during the latter 50% of reaction is almost identical under all three reaction conditions shown. Thus, the slow population reacts similarly, regardless of ionic environment. Given the fact that the rate transition occurs at the same extent of reaction but at different times under the different ionic conditions (~10 minutes for $NH_4SO_4$; ~30 minutes for $NH_4Cl$,

~60 minutes for KCl), the two conformations of molecules appear to be in slow exchange, if they exchange at all. This finding indicates that a kinetic treatment of group II intron self-splicing in monovalent ions must accomplish two objectives: It must account for the two chemically distinct pathways for splicing (branching and hydrolysis). At the same time, it must also mathematically account for two non-interconverting conformations in the precursor RNA.

**Kinetic treatment of two RNA conformations:** The preceding data suggest that biphasic reaction behavior is due to the presence of two populations of molecules, irrespective of the type of reaction pathway that dominates during splicing. In addition, the data suggest that biphasic behavior should not be attributed to a sequential pathway in which lariat serves as an intermediate for production of linear intron. This is because self-splicing proceeds almost exclusively through a hydrolytic pathway in the presence of KCl, although the same biphasic kinetics are observed under that reaction condition. The most simple kinetic model for the results is presented in Scheme 2, in which the precursor consists of two non-interconvertible populations (A and B) - each of which is catalytically active. Each of these populations may be able to splice through the hydrolysis or branching pathways. Each population may therefore react with the same parallel-first-order kinetic behavior outlined in previous sections.




Scheme 2

The presence of two reactive non-interconverting populations requires that the raw data in Fig. 26A-C fit two sets of exponential terms instead of one, as shown previously.

For this purpose, equations 1-5 describing parallel first-order behavior were modified such that product evolution is described by an equation containing two sets of terms: one describing population A and one describing population B (equations 6-10). As before, if parallel kinetics is followed, $k_T$ values for a single population (k, for example) will be the same, regardless of whether linear, lariat or precursor data is fit. Fractional values for a given population will remain a constant throughout reaction, and the four fractional coefficients describing extent of formation of intronic products will add up to one.

**Self-splicing in 100 mM $Mg^{2+}$ and 500 mM KCl:** The data obtained under this condition was best fit using biexponential equations 6-8, as predicted from the first order plots (Fig. 25C). Both the linear and lariat curves fit biexponential equations 7 and 8, rather than a sequential model (Tinoco et al., 1978), suggesting that the pathway outlined in scheme 2 is correct. Both populations of molecules partitioned between linear and lariat reactions (Fig. 26A). As expected for a parallel first order reactions from each population i, all kvalues matched and all four fractional values describing each population added to one (legend, Fig. 26A). It is important to note that even the fast population of molecules (totaling 41% of the population in this experiment) reacted both through the linear and lariat pathways in competition with one another. The apparent rate of the reaction by population A was 0.028 $min^{-1}$ (note that error limits for all rates are provided in the caption to Fig. 26). By solving equations 9 and 10 for population A, we observe that hydrolysis is the fastest process, with an individual rate k= 0.022 $min^{-1}$, compared to k= 0.0063 $min^{-1}$ for branching (Table 5, legend, Fig. 26A). For the slow conformation (population B), hydrolysis is also faster than branching. The overall rate $k_T$ = 0.0042 $min^{-1}$ and, by solving equations 9 and 10, we observe that the k= 0.0032 $min^{-1}$ and k= 0.0011 $min^{-1}$. These latter rate constants

for B are remarkably similar those observed for splicing of the entire precursor population in 100 mM $MgCl_2$ alone (Table 5).

**Kinetics of self-splicing in $NH_4Cl$:** Another monovalent salt that is a standard condition for studies of in-vitro splicing is $NH_4Cl$ (Peebles et al., 1987; Muller et al., 1988; Jarrell et al., 1988b; Suchy & Schmelzer, 1991; Winkler & Kuck, 1991; Peebles et al., 1993). Time-courses of reaction in 100 mM $MgCl_2$ and 500 mM $NH_4Cl$ are shown in Fig. 26B. There are several trends that are apparent from inspection of the raw data. First of all, the lariat molecules are formed during the first 50% of reaction, after which there is a plateau in lariat production. Large amounts of linear intron are observed only during the second half of reaction. A simplistic interpretation of these results might be that the linear band is due to lariat decomposition. But, after lariat begins to plateau (~50 minutes), all of the evolution of linear intron is due to decrease in precursor (Fig. 27B). Once lariat evolution is a flat line, the curves for precursor and linear intron change inversely with respect to one another. It is clear that most of the lariat that is produced over time is stable because lariat molecules remain the dominant product after 15 hours of reaction (Fig. 27, lane 5). Therefore, even by this qualitative analysis, it appears that linear intron is produced independent of lariat, through a mechanism in which a hydrolytic pathway competes with branching.

Decrease in precursor and evolution of lariat are best fit by biexponential equations 6 and 8, suggesting that lariat is produced from two conformations of precursor molecules. However, the evolution of linear intron fits only a single exponential (equation 2). The lack of a biexponential fit for the linear intron suggests that linear intron is produced only from the slow population (B). That B produces both linear and lariat intron through parallel first-order

kinetics is supported by the fact that $k_T$ obtained from equation 2 is the same as k, obtained from the fit of the lariat data to equation 8. Taken together, these data show that fast population (A) of precursor reacts solely through a lariat pathway, with a rate of 0.038 $min^{-1}$. In ammonium salts, the hydrolytic pathway from conformation A is completely shut down. Under these conditions, this conformer may be able to exclude water and promote an exclusively branch-mediated splicing pathway. Conversely, the slow population (B) obeys parallel first order kinetics, partitioning through competing linear and lariat pathways. Under these reaction conditions, Scheme 2 must be slightly modified to a form in which there is no arrow connecting population A with the linear intron (the kpathway). Taking the precursor depletion as the most accurate representation of overall reaction for k, and using the relative amounts of linear and lariat produced by that (f= 0.51, f= 0.10), as coefficients for correcting the data, the rates of splicing from B were calculated to be k= 0.0032 $min^{-1}$ and k= 0.00063 $min^{-1}$. These rates of reaction by population B are remarkably similar to those observed under the other reaction conditions (Table 5).

Kinetics of self-splicing in $(NH_4)_2SO_4$: Like $NH_4Cl$, lariat formation is radically stimulated by $(NH_4)_2SO_4$ (Fig. 26C). It appears that population A splices only through a branching pathway, resulting only in lariat intron. One salient difference in the kinetic behavior in $(NH_4)_2SO_4$ salt is the existence of a considerable burst early in the reaction. When plotted in a semi-log format, in addition to the sharp demarcation observed at 50% reaction (between fast and slow populations of molecules), there is a large burst consuming 13% of population A during the first 60 seconds of reaction (inset, Fig. 25A). We observe that, as in the case of $NH_4Cl$, the linear species evolves very slowly and fits a single exponential. After a plateau in lariat formation, linear evolves only at the expense of precursor (Fig. 26C).

The timecourses in $(NH_4)_2SO_4$ were fit using methods similar to that described for $NH_4Cl$, with one exception designed to account for the burst of reaction that immediately consumes 13% of the precursor (Fig. 25A, inset). This fractional amount was used to add an additional rate term (for which the f = 13%) to equations 6 and 8 so that the remainder of the population could be fit precisely. The equations were extended in the same way as equations 6 and 8 were adapted from equations 1 and 3, although the new f value was held constant. This served to reduce the uncertainty of adding an additional set of variables to the fit. After fitting the precursor and lariat data to the modified equations 6-8, it was possible to solve for individual rates constants for branching by population A. The burst population (contributing a fraction of 0.13) reacted with a rate of 0.22 $min^{-1}$. The remainder of population A (f) reacted solely through a branching pathway, with a rate of 0.047 $min^{-1}$ (Table 5, Fig. 26C). The slow population B ($k_T$ = 0.0028 $min^{-1}$) reacted through both hydrolytic (f= 0.39) and branching (f= 0.22) pathways at respective rates of k= 0.0018 and k= 0.0010 $min^{-1}$ (Table 5). Population B, therefore shows competitive behavior and rate constants similar to the other reaction conditions. Population A, however, is quite different in that it undergoes a rapid burst of lariat formation that is among the fastest rates seen for reactions by group II intron. This confirms earlier observations indicating that, if it weren't for conformational states that obscure it, the chemical rate of branching would appear to be extremely rapid (Chin and Pyle, 1995). The value of the burst provided here (0.22 $min^{-1}$), may be viewed as a lower limit for the chemical rate of branching.

The time-courses of splicing under different reaction conditions display mechanistic features that argue against a sequential pathway: (a) There is no "hump" in lariat

formation, which would appear as an initial burst followed by a steady decrease in the lariat population (Fersht, 1985). Such a rise and fall has been observed before under conditions where the branch is destabilized (Chin and Pyle, 1995). (b) The data fit equations for parallel rather than sequential kinetic models. (c) Even when linear intron dominates, lariat molecules continue to evolve throughout reaction, arguing that their conversion to linear molecules is not so rapid that lariats cannot accumulate. (d) Linear appears to evolve quantitatively at the expense of precursor, rather than at the expense of lariat.

**Ribonuclease analysis of the linear intron species:**

Large RNA molecules, whether linear or lariat are inevitably susceptible to some degree of degradation at long times. One cut in a lariat molecule will linearize it, creating a species that migrates with hydrolytically generated intron. The branched "tail" protruding from the linearized lariats is very small, so it is expected to have only minor effects on mobility (Peebles et al., 1986). Although the data presented here are consistent with a hydrolytic pathway for production of linear intron, a fraction of the linear species is expected to result from lariat decomposition. This is supported by the kinetic data, which show that all $k_T$ values determined from lariat evolution are slightly larger than the $k_T$ values determined from linear intron evolution. This effect stems from an artificially early plateau in the evolution of lariat, giving rise to an artificially late plateau in linear intron production (due to slow conversion of lariat to linear). This results in a shorter $t_{1/2}$ for lariat and a longer $t_{1/2}$ for linear intron evolution. Because we did not expect the linear intron to be purely the result of a hydrolytic pathway, we were interested in determining the relative stoichoimetry of branched to mononucleotides in the linear band. We reasoned that linear intron isolated after long times of reaction might contain more broken lariat than linear species

isolated after short times of reaction.

In order to obtain direct physical information on the branched trinucleotide content of the linear intron, we subjected the linear intron species to chemical characterization, through digestion with P1 nuclease . Lariat intron contains a single branched trinucleotide, of composition pA(2'-pG)3'-pU. P1 nuclease is unable to cleave the phosphodiester linkages within a branched trinucleotide (Reilly et al., 1989). Therefore, complete digestion of the lariat RNA with P1 nuclease results in abundant mononucleotides and a single branched trinucleotide. These two species are readily separable by PAGE and can be quantitated by radioanalytic scanning if one of the components of the branch is labelled with $^{32}P$ (Moore & Sharp, 1992; Chin and Pyle, 1995). When linear intron results from lariat breakage, it will contain a branched, 2'-5' linkage and P1 digestion will result in a particular stoichiometry of mono and trinucleotides.

Of the three nucleotides contained in the branched species (pA(2'-pG)3'-pU), guanosine is the least abundant in the rest of the intron. Therefore, we transcribed precursor RNA with $\alpha$-$^{32}P$-GTP and subjected this RNA to three diverse conditions that promote efficient splicing: 100 mM $MgCl_2$ with 500 mM KCl, $NH_4Cl$ or $(NH_4)_2SO_4$. Large scale aliquots were removed at 50% and 90% extents of total reaction and products were separated by preparative-scale PAGE. Lariat and linear intron bands were isolated from the gel, and the RNA was digested to completion with P1 nuclease. After product separation, the nucleotide ratio in the lariat band was approximately 120:1 under all conditions (Table 6). There are 119 guanosines in the intron and therefore P1 digestion of the lariat RNA should result in an ideal mono to trinucleotide ratio of 118:1, which is very close to the observed value. This confirmed the precision of analysis and suggested that 2'-5' branched linkage was stable during

isolation from the gel.

**Table 6: P1 ribonuclease analysis of intronic phosphodiester linkages:**

| Condition | Lariat Band | Linear Band | Fraction of linear intron |
|---|---|---|---|
| (buffer salt[a] & minutes reaction) | mono/branch nucleotides | mono/branch nucleotides | containing a branch[c] (for whole molecules) |
| | | | |
| $(NH_4)_2SO_4$ | | | |
| 30 | 121[b] | 615[b] | 19.7±1.5% |
| 180 | 118 | 254 | 46.4±3.5% |
| | | | |
| $NH_4Cl$ | | | |
| 30 | 112 | 667 | 16.8±1.3% |
| 180 | 114 | 237 | 48.1±3.6% |
| | | | |
| KCl | | | |
| 40 | 105 | 523 | 20.1±1.5% |
| 180 | 119 | 297 | 40.1±3.0% |

**Table 6 Legend:** Splicing reactions were conducted under the conditions indicated, followed by electrophoresis and isolation of individual intronic products (RNA was internally labelled by incorporation with $^{32}$P-GTP). The separate lariat and linear products were digested to completion with P1 nuclease and relative proportions of mononucleotide to branched trinucleotide in each species were compared. There are 118 guanosine residues in the intron, one of which becomes incorporated in the branched trinucleotide upon adenosine attack at the 5'-splice site. A ratio of mono/branched nucleotide that is significantly greater than 118 indicates formation of linear intron

through an event that does not involve branching (i.e. hydrolysis during the first step of splicing). [a]Salt indicated was present at 0.5 M, together with 0.1 M $MgCl_2$. [b]Variance in each nucleotide ratio is ±5.8% with 95% confidence, representing error in lariat values show representing six different trials. [c]Error is calculated with 95% confidence for four separate trials, represent error between values obtained in separate experiments performed on different days.

P1 analysis of the linear intron band showed that, during the first half of reaction, only 20% of the linear species contained a branched trinucleotide (Table 6). Random RNA degradation should increase with time and therefore it was notable that at 180 minutes, after the latter half of reaction in each case, the amount of linear species that arose from lariat molecules was approximately 45% (Table 6). The late timepoints are important because they are an internal control, demonstrating that we would observe branched trinucleotide if it were there. These timepoints also show that, when significant amount of RNA degradation is allowed to occur, one will observe that the linear band contains a significant fraction of branched trinucleotide. Nonetheless, in our system, the actual amount of branched species in the linear product was never dominant, even at 180 minutes. Thus, the nucleotide ratio obtained after P1 digestion of the linear band indicated that the major component was true linear intron.

**Spliced-exon reopening is unrelated to hydrolytic splicing:** During in-vitro self-splicing of the ai5g intron, one commonly observes free 5′ and 3′ exons (Fig. 23, 27). It has been proposed that the free exons arise from a secondary reaction involving rapid hydrolysis of spliced exons (spliced-exon reopening, or SER) (Peebles et al., 1987; Jarrell et al., 1988b). This latter model is credible given that the ratio of 5′ to 3′-exons is always one (data not

shown) and it was unlikely that hydrolysis at both splice sites would take place at identical rates. In addition, timecourses of SER reveal that spliced exons predominate early in reaction, giving way over time to free exons (Fig. 27). By comparing the amount of spliced to linearized exons at various points during the reaction, it is possible to discern the relative efficiency of the SER reaction under different reaction conditions (Table 7).

**Table 7: Dependence of the SER reaction on ionic conditions.**

| Ionic Content[a] of Reaction Buffer | Fraction of Ligated Exons[b] at 12% Reaction | Fraction of Ligated Exons at 25% Reaction |
|---|---|---|
| | | |
| 0.5 M $NH_4Cl$ | 0.21±0.04[c] | 0.11±0.03 |
| 0.1 M $MgCl_2$ | | |
| | | |
| 0.5 M $(NH_4)_2SO_4$ | 0.63±0.13 | 0.50±0.11 |
| 0.1 M $MgCl_2$ | | |
| | | |
| 0.5 M KCl | 0.071±0.01 | <0.01[d] |
| 0.1 M $MgCl_2$ | | |
| | | |
| 0.1 M $MgCl_2$ | 0.027±0.005 | 0.033±0.007 |
| | | |
| 5 nM protamine | 0.56±0.11 | 0.36±0.08 |
| 0.01 M $MgCl_2$ | | |

**Table 7 Legend:**

a. Other than the specific salts indicated, all reactions contained 40 mM MOPS, pH 7.5 and were conducted as described in the experimental section.

b. This value is the fraction of ligated exons/total

released exons. Total released exons = the ligated exons plus the free exons produced through the spliced-exon (ligated-exon) reopening reaction (SER). That free exons were formerly ligated and cleaved through SER is indicated by previous work, by the higher fraction of spliced exons early in reaction (see above), and by the fact that the ration of fee 5'-exon is always equal to 1.

c. Variances shown are the standard errors calculated using a 95% confidence level from the mean squared deviation in six to eight trials under each condition.

d. Spliced exons were present at levels that approach the limit of detection.

Because conditions that promote hydrolytic splicing are sometimes accompanied by SER (Jarrell et al., 1988b), one might assume that a hydrolytic first step is accompanied by an unproductive side reaction that reopens spliced exons. If this were true, then a hydrolytic first step might be considered detrimental to productive splicing and branching would be considered the "proper" pathway for self-splicing.

However, we observe SER under all reaction conditions, regardless of whether branching or hydrolysis dominates during the first step. Even reactions containing high concentrations of ammonium are accompanied by SER. Particularly $NH_4Cl$, which promotes splicing primarily through lariat formation, also promotes rapid SER (Fig. 27, Table 7). The amount of SER determined in this case indicates that spliced exons are hydrolyzed almost as fast as they form. The same is true under conditions of 100 mM $MgCl_2$, without any monovalent ion. Conversely, protamine stimulates a predominantly hydrolytic splicing pathway that leads primarily to spliced exons (Fig. 27, Table 7). An obligate relationship between SER and the hydrolytic first step of splicing is not evident when a variety of conditions are considered together. The prevalence of SER is instead correlated with the presence of chloride ion, which is often provided in molar amounts in the study of self-splicing.

However chloride ion is not singularly responsible for SER because, although it is less prevalent, the reaction is still observed in $(NH_4)_2SO_4$ (Fig. 23A) and in reactions where chloride has been completely eliminated (K(acetate)/Mg(acetate)$_2$, data not shown).

DISCUSSION:

**The existence of two possible pathways for the first step of self-splicing:**

The evidence for a competitive interplay of two different reactions at the 5'-splice site is based on a variety of results, including the kinetic behavior observed during timecourses of group II intron self-splicing (Fig. 24 and 26). This is supplemented by physical characterization of reaction products (Table 6). The kinetic analysis indicates that the decrease in precursor RNA and the evolution of linear and lariat products follow a kinetic paradigm known as the "parallel" reaction (Jencks, 1987). As such, the individual products arise simultaneously from a common precursor. More qualitatively, it is observed that the lariat intron is too stable to serve as the precursor for the majority of linearized intron observed and linear intron is observed to evolve at the expense of precursor, rather than lariat RNA. Physical analysis on the products of reactions that favor both hydrolytic and branching pathways indicates that the majority of linear molecules do not contain a branched trinucleotide (Table 6). Particularly early in reaction, before RNA degradation sets in, linear intron is almost entirely composed of molecules arising through the hydrolytic pathway. Therefore this species is primarily the product of a different splicing pathway should not necessarily be considered a "broken lariat".

The hydrolysis pathway may be a legitimate mechanism for the first step of group II intron self-splicing. After all, RNase P catalyzes phosphodiester cleavage using water or hydroxide as the nucleophile (Guerrier-Takada et al., 1983;

Pace & Smith, 1990). Regardless of any hypothetical relationship to RNase P, it is possible that components of the group II intron active site have evolved specifically to activate a water/hydroxide nucleophile. There is little reason to assume that hydrolytic attack takes place in a different active site than branching, particularly since branch-mediated and hydrolytic attack at the 5'-splice site both proceed with identical preference for an Sp phosphorothioate diastereomer in cis-exon constructs (Podar et al., 1995a). Rather than invoking an active site that is specifically designed to activate a water nucleophile, one could alternatively argue that the "normal" nucleophile is the 2'-hydroxyl group of the branch-point A in Domain 6. If this is the case, hydrolysis may occur because the active site exerts tremendous strain on, or creates significant chemical activation of, the scissile bond at the 5'-cleavage site. As a result, other nucleophiles in the vicinity (particularly 55M water) may cleave the scissile phosphodiester linkage.

**Implications of a hydrolytic pathway:**

A general hydrolytic mechanism for group II intron self-splicing has several important implications. The most important one is that, if this reaction readily occurs in-vitro, it may have a function in-vivo. Unlike the facile reversibility of branch-point attack (lariat formation), hydrolytic scission of the 5'-splice site is effectively irreversible (Chin and Pyle, 1995). This has important implications for intron mobility, which can occur through lariat reinsertion into sites resembling the original 5'-exon sequences (Lambowitz and Belfort, 1993). It is generally believed that group II introns are infectious elements that came to rest in their present locations as a result of reactions that resembled reverse-splicing (Lambowitz, 1989). A possible role for hydrolytic splicing comes from the fact that, once a group II intron has become integrated into a cascade of normal gene expression, the

cell may try to prevent the accumulation of highly reactive lariat introns. Lariat introns are potentially deleterious because they can reverse-splice (Augustin et al., 1990; MŸller et al., 1991), reintegrating into other RNA molecules. For certain introns with a high propensity to migrate, there may have been selective pressure for the cell to develop intron mutations that specifically promote hydrolytic splicing. In that way, the cell exploits the advantages of splicing, without the risk of further genomic alteration.

Another possible role for a hydrolytic splicing pathway is related to the fact that many group II introns contain open reading frames for the translation of proteins called "maturases" (Lambowitz and Perlman, 1990; Lambowitz and Belfort, 1993). These proteins aid in splicing and mobility of group II introns. If translation necessarily occurs after splicing (and it is not clear that this is the case), then there might be serious difficulties in using lariat RNA as the message. Although translation in mitochondria is not well understood, potential problems with lariat translation are suggested by recent studies showing that cytoplasmic translation initiation of circular RNA molecules can only occur if these contain particular sequences called internal ribosome entry sites, or IRES (Chen & Sarnow, 1995). Even if the hydrolytic pathway were very inefficient and rare in-vivo, it might serve an important function: Hydrolytic splicing might generate a small population of linear intron molecules that are readily translated. This would result in the accumulation of maturase proteins, which would then facilitate additional rounds of self-splicing through a predominantly lariat pathway. By this model, linear introns may serve to "seed" lariat-mediated splicing by generating the proteins required for rapid self-splicing in-vivo.

A hydrolytic splicing pathway in-vivo is consistent with reaction conditions that favor hydrolytic splicing in-vitro.

Intracellular salt concentrations are approximately 20 mM $MgCl_2$ and 140 mM KCl (Alberts et al., 1989) which, together with millimolar levels of polyamines and cationic peptides, are most analogous to the in-vitro conditions that favor the hydrolytic splicing pathway. Indeed, the in-vitro conditions that promote lariat formation (molar ammonium ion) are the least physiological. Nonetheless, group II introns were first identified because of an abundance of stable circular RNA molecules observed during electron microscopic analyses of cellular RNA (Arnberg et al., 1980; Halbreich et al., 1980). There is good evidence that splicing in-vivo proceeds through a lariat pathway, although there is no evidence suggesting that this is the exclusive route or that the linear pathway does not occur. Indeed, lariat molecules are expected to appear more abundant because they do not have free ends and are therefore refractory to attack by cellular nucleases.

Electron microscopy (Arnberg et al., 1980; Halbreich et al., 1980; Hensgens et al., 1982; Schmidt et al., 1987), combined with enzymatic analysis of cellular RNA (van der Veen et al., 1986), showed that excised group II introns were lariats, containing an internal 2'-5' bond and a free 3' terminus. However, abundant linear RNA molecules of the same contour length were also identified in the RNA fractions used for electron microscopy (Arnberg et al., 1980; Schmidt et al., 1987). In group II intron i1 from the mitochondria of filamentous fungus *Podospora anserina*, two forms of the intron were isolated (Schmidt et al., 1987). Mapping of the intron with different DNA probes indicated that these two intron forms contained the same sequences. RNAse H digestion converted the slower migrating form into one that comigrated with the faster form, indicating that the RNA transcripts were topological isomers of the same molecule. Debranching treatment also converted the slower band into one which comigrated with the faster band, indicating that the RNA in the slower band is a lariat form of the RNA (Schmidt et al., 1987). It is interesting to note that only one band results from the RNase

H treatment, and it is not accompanied by a smear that would be expected if the linear species contained randomly broken lariat molecules. The true test of a hydrolytic pathway in-vivo does not, however, rest with physical analysis of cellular RNA because alternative explanations (such as instability of the 2'-5' link) can found for the results. Instead, evidence for the existence of this pathway lies with genetic studies on mutant yeast strains in which the branch-point A has been deleted from essential group II introns like ai5g. These studies, in which the bulged A has been inactivated in a variety of ways, would be the most definitive way to unravel the function of hydrolysis in-vivo.

Four group II introns contained within chloroplast $tRNA^{Val}$ genes from four diverse organisms (*Nicotiana tobacum*, *Marchantia polymorpha*, *Zea mays*, and *Hordeum vulgare*) are known to be missing the branchpoint A in Domain 6 (Michel et al., 1989). These four introns might be good candidates for a studying a class of group II intron that splices entirely through a linear pathway in vivo. Not all group II introns may be alike in their ability to splice hydrolytically in-vivo, so it might be important to test branch-point deletion on a variety of essential yeast group II introns. It is also possible that the different pathways are used under different growth conditions, or at different times during organellar biogenesis.

**Mechanistic features of the self-splicing reactions:**

One of the most pronounced effects explored in this study is the presence of two nonexchangable conformations in the precursor RNA. Population A (fast) can splice through either hydrolytic or branching pathways, although it dramatically prefers the branching pathway in the presence of ammonium ions. "A" forms only in the presence of high monovalent ion, and it appears to be a more reactive form of the intron. Population B (slow) can also utilize either pathway, although it prefers hydrolysis. "B" appears to be some sort of default form of the intron, with a similar kinetic profile under all reaction

conditions including low salt. Indeed, under low salt conditions, this may be the only conformation of the intron. Any future analysis of splicing rates will have to contend kinetically with these multiple conformations. Our means of dealing with this was to separate the conformations mathematically using multiple exponentials in our rate equations. Preliminary gel-filtration results (data not shown) suggest that it may be possible to physically separate the conformations and study them individually.

Conformations A and B might be related to the incubation procedures used in this study. Unlike any other group II intron splicing work published to date, the reactions shown here were preceded by a step in which the RNA was heated to 95¡C, in order to melt alternative inactive conformations formed during RNA storage (Walstrum & Uhlenbeck, 1990; Pyle and Green, 1994; Uhlenbeck, 1995). In order to determine if the conformational distribution was related to pre-heating, we conducted self-splicing reactions on precursor RNA that was not subjected to 95¡C preincubation (data not shown). We observed behavior very similar to that reported here with one important difference: Population A reacted with a rate ~50% slower than that observed under the other conditions reported here. However, the same relative proportions of A and B were observed. This suggests that a short 95¡C preincubation does not adjust the relative proportion of A and B (or allow their interconversion), but it aids in activation or folding of population A. The observation of two conformers is consistent with a recent report from another laboratory in which, under different preincubation conditions than reported here, 40% of the ai5g intron was found to be less reactive (Podar et al., 1995b).

The rate constants reported in Table 5 are useful benchmarks for structure/function studies on mutants of group II introns. However, more experiments would be required to determine if these rate constant represent specific events in the first step

of splicing, such as chemistry or conformational change. It may be possible to draw parallels with mechanistic analyses performed on multi-component ribozyme constructs of the ai5g group II intron. The fast rate constant for hydrolytic splicing in KCl (k) is virtually identical to the chemical rate of Domain 5-catalyzed hydrolytic scission of the 5'-exon from Domain 1-3 (Pyle and Green, 1994) under the same ionic conditions. In a study that analyzed trans-D56 attack on the 5'-splice site (Chin and Pyle, 1995), it was observed that branching was limited by the rate of conformational sampling of a chemically active state. This rate was analogous to kunder the reaction conditions reported in Table 5. Similarly, a burst in reactivity was proposed to represent a lower limit for a chemical step of branching, which is necessarily very rapid (Chin and Pyle, 1995).

Although the reaction condition involving protamine had not been reported in previous studies of group II intron self-splicing, other charge-neutralizing proteins have been found to replace the high monovalent ion requirements of certain ribozymes. For example, RNase P can function efficiently at lower salt concentrations in the presence of the C5 RNase P protein (Reich et al., 1988). While there may be specific proteins that associate with particular group II introns in-vivo (Wiesenberger et al., 1992; Van Dyck et al., 1995), we were interested in examining the effects of a common nonspecific RNA and DNA-binding protein. The amino acid sequences of protamines (short arginine-rich peptides) resemble isolated RNA-binding domains of about 30 amino-acids in length. Structural studies of protamine binding to RNA show that the polypeptide forms an α-helix, which rests in accessible helical grooves of nucleic acid duplexes (Warrant & Kim, 1978; Hud et al., 1994). Our first experiments with protamine were conducted at μM concentrations of the peptide. After decreasing the [protamine] such that there were stoichiometric quantities of protamine bound to the RNA, we began to see stimulation of catalysis at 10 mM $MgCl_2$. Protamine had the

greatest stimulatory effect at a concentration of 5 nM (the precursor was generally present at 1 nM). In addition to stimulating the reaction in the presence of 10 mM $MgCl_2$ (Fig. 24B,C,D), gels of the reaction products showed that the RNA was not even slightly degraded during these reactions (Fig. 27, lanes 18-20). Despite the fact that most of the reaction proceeds through a hydrolytic pathway during the first step of splicing, protamine promotes a high level of exon ligation (Fig. 27, lanes 18-20, Table 7), without a concommitant SER side reaction. Thus, protamine seems to stimulate a stable conformation of the intron at low salt and, while it is not the most reactive ionic condition analyzed, it promotes clean self-splicing without side reactions.

**Relationship of results to previous studies:**

Self-splicing by group II introns was first evident in reactions showing the evolution of two intronic product bands (Peebles et al., 1986; Schmelzer and Schweyen, 1986; van der Veen et al., 1986). The slower migrating intronic product was shown to be a lariat molecule. Initial studies of the fast-migrating linear species concluded that, except in KCl, it consisted primarily of "broken lariat" (Peebles et al., 1986; Peebles et al., 1987; Jarrell et al., 1988b). This is a reasonable conclusion because even a single break in lariat intron would result in an RNA with the same mobility as linear intron. However, a hydrolytic pathway for production of linear intron was never discounted and several pieces of evidence indicated that linear intron might be significant reaction product: First of all, the "broken lariat" band was refractory to debranching treatment by HeLa cell S100 extract (Peebles et al., 1986). Second, the lariat product, while easily degraded during the isolation procedure, was shown to be stable to reincubation under splicing conditions (Peebles et al., 1986), indicating that any conversion of lariat band to "broken lariat" was a process too slow to explain the amount of linear intron generated under the reaction conditions. This result also indicated that linear species arising solely by

"self-debranching" is also unlikely, a result that has been confirmed by later work on the relative stability of the 2'-5' linkage (Chin and Pyle, 1995).

Additional biochemical data supported the notion that the fast-migrating intronic species contained significant quantities linear intron. The molecular nature of the linear intron/broken lariat band was investigated by T1 ribonuclease fingerprinting and by branchpoint trinucleotide analysis with nuclease P1 (Peebles et al., 1987). Data from these studies indicates that a fraction of the "broken lariat" band consisted of linear species. Quantitation of the branchpoint trinucleotides revealed that approximately 80% of the linear/broken lariat band seen in KCl reactions was linear intron, while 20% was broken lariat. In ammonium sulfate, the raw data suggest that approximately 50% of this band was linear intron after 30 minutes (Peebles et al., 1987). Jarrell et al. concluded that hydrolytic cleavage at the 5' splice site is a bona fide pathway utilized by the group II intron under KCl conditions in vitro (Jarrell et al., 1988b).

That a hydrolytic pathway could mediate self-splicing was particularly evident from experiments studying the role of the branchpoint A, which showed that a bulged adenosine is dispensible for self-splicing in vitro (van der Veen et al., 1987). Self splicing can occur under $(NH_4)_2SO_4$ conditions in vitro when the branchpoint A is either deleted or basepaired in a helix. In both cases complete splicing proceeds (albeit more slowly) via an apparently linear pathway, since only a linear intron product was seen and no lariat band could be identified. The possibility that rapid cleavage converted all lariat product into "broken lariat" was discounted by isolating the putative linear band, digesting with T2 ribonuclease, and isolating the trinucleotide on two-dimensional TLC (van der Veen et al., 1987). No branchpoint trinucleotide was recovered in the basepaired mutant. Splicing by an intron without a functional branch-point has also been observed in our

laboratory during mutational studies on the bulged adenosine. When the branch-point A is replaced by a C, lariat molecules cannot be detected. However, hydrolytic splicing (in both 0.5 M $(NH_4)_2SO_4$/100 mM $Mg^{2+}$ and 100 mM $Mg^{2+}$ conditions) proceeds with the same rates as reported here for the wild-type intron (Qiaolian Liu and A.M.P., manuscript in preparation). This provides further evidence for a linear pathway and confirms that, using the parallel kinetic mechanism on the wild-type intron, hydrolytic rate constants were correctly deconvoluted from branching rate constants in the present study (Table 5).

This study does not contradict previously published data, but extends understanding of the first step of splicing in several ways. The data indicate that a hydrolytic pathway for splicing competes directly with branching at all times, under all conditions tested. Certain salts or proteins can preferentially promote one pathway or the other, and either route can be fast. Parallel reaction kinetics are followed, and linear intron should not necessarily be considered a secondary or degradative product. Hydrolytic cleavage in the first step of splicing does not prevent the second step, although rate reductions in the second step cannot be ruled out since the first step remains rate-limiting under all conditions. We find that hydrolytic splicing is no more likely than branch-mediated splicing to be accompanied by the destructive SER reaction. The breadth of conditions analyzed here indicate that the hydrolytic pathway is not an artifact of high salt reaction conditions, because the linear pathway can proceed, and even dominate, under low salt conditions similar to physiological. Given that a group II intron has the chemical capability to react through two distinctive pathways, and that the reaction environment regulates the choice of dominant pathway, it is important to consider the potential involvement of either pathway in the many important biological functions of group II introns.

MATERIALS AND METHODS:

**Plasmids and transcripts:** The ai5g intron and flanking exons from *Saccharomyces cerevisiae* were cloned downstream from a T7 RNA polymerase promoter, and are contained in plasmid pJD20 (Jarrell et al., 1988a). Unspliced precursor RNA was transcribed from plasmid pJD20 after linearization by restriction enzyme Hind III. The precursor RNA is 1497 nucleotides in length, consisting of 887nt of intron flanked by 292 nucleotides of 5'-exon and 318 nucleotides of 3' -exon. Using methods described previously (Pyle and Green, 1994), transcriptions were performed in the presence of $\alpha$-$^{32}$P-UTP, resulting in precursor RNA that was internally labelled throughout the molecule. Although a small amount of RNA splicing takes place during transcription (the buffer contains 10 mM $Mg^{2+}$), the major product was precursor RNA. Transcription was stopped after 1.5 hours by adding an equal volume (usually 20ul) of gel loading dye (10M urea, 0.1% xylene cyanol and bromophenol blue dyes, 40mM tris pH 7.5, 8.3% sucrose, and 0.83mM EDTA, final concentrations) and then loading directly on a 4% denaturing polyacrylamide gel. Following electrophoresis, the band was visualized by autoradiography for 2 minutes, excised from the gel, and then eluted overnight in MOPS storage buffer (MOPS SB: 10mM MOPS, 1mM EDTA, pH 6.5). RNA was then isolated by ethanol precipitation of the eluate (3 volumes of 100% ethanol, 0.3M sodium chloride) and then washed with one volume of 70% ethanol. The RNA was dried, resuspended in 20ul of MOPS SB, and stored at -20°C. RNA concentrations were determined from their specific activities.

**Splicing Reactions:** The splicing reactions were performed using a total volume of 20 to 40 $\mu$l, in 0.65 ml Slickseal tubes. The concentration of precursor RNA for the experiments reported here was 1 nM, although all reactions were repeated at 0.1 nM precursor to observe any concentration effects that might result from a sequential, rather than parallel reaction pathway. No differences in rate were observed. Reactions were initiated by preincubating the precursor RNA (in MOPS SB) at 95¡C for one minute. The RNA sample was then allowed to cool

to reaction temperature before combining with an equal volume of a salt mix, such that the final ionic reaction conditions were as indicated (text, Figure 26, Table 5). The various ionic conditions studied were 10 mM $MgCl_2$ (Aldrich, 99.99%) with 2 mM spermidine or 5 nM protamine (Sigma); 100 mM $MgCl_2$; and 100 mM $MgCl_2$ with 500 mM KCl (99.99% pure, Aldrich), 500 mM $NH_4Cl$ or $(NH4)_2SO_4$. Before use in splicing reactions, protamine was purified by gel-filtration chromatography. Regardless of ionic condition, each sample was buffered with 40 mM MOPS, pH 7.5 and the reactions were incubated at 42¡C. To measure extent of reaction at specific timepoints, aliquots (1 or 2 ul) were removed from the reaction mixture at the specified time intervals. The aliquots were placed in equal volume of quench buffer (1.8% sucrose, 1X TBE, 0.018% xylene cyanol dye, 36% formamide, and 25mM EDTA) and placed on ice to stop the reaction. Samples were then loaded onto a 4% denaturing polyacrylamide gel and products were resolved by electrophoresis. The gels were dried and quantitated on a Packard Instantimager, a radioanalytic scanner.

For the low-salt splicing reactions conducted solely in 100 mM $Mg^{2+}$, or in 10 mM $Mg^{2+}$ with spermidine or protamine, the type of base used in pH adjustment of the MOPS buffer had significant effects on the reaction. Under the above conditions, when pH of the reaction buffer was adjusted with either KOH or $NH_4OH$, reaction products were indistinguishable and the effects described in Fig. 3 and in the text were observed. However, when pH was adjusted with NaOH, hydrolytic cleavage at the 5'-splice site was inhibited several-fold (provided no other monovalent salts were added). Because pH adjustment introduces small quantities of particular monovalent cations that may have inhibitory or stimulatory effects, it is important to consider the effects of additional components in the reaction buffer while measuring splicing at low salt.

**Quantitative analysis of splicing reactions:** Radioactivity in individual bands was quantitated on the radioanalytic scanner.

Counts from a nearby blank position on the same lane were subtracted (after normalizing for decreasing counts/mobility in the t = 0 lane) as an internal measure of background RNA degradation at each timepoint. This was facilitated by running separation gels (4% acrylamide) extensively, in order to cleanly separate all products. The relative fractions of precursor, lariat, linear, and exonic products were plotted against time and fit to equations 1-3 or 6-8, as shown in the text using Kaleidagraph (Abelbeck Software). The reaction of unspliced precursor RNA and evolution of both of its two products (lariat and linear intron) were each fit individually. Under all reaction conditions, self-splicing proceeded to completion (Fig. 27A, lane 15), thus eliminating the need for endpoint correction of the data. Equations used for fitting the data are presented as follows:

For a single population of molecules reacting through parallel kinetics, the independent evolution of two products (lariat and linear intron) is described by the following equations, which were applied in analysis of data from the 100 mM $Mg^{2+}$ and 10 mM $Mg^{2+}$, 6 nM protamine reaction conditions (Fig. 24A, B).

$$\frac{\text{Precursor}(t)}{\text{Precursor}_0} = e^{-k_T t} \quad (1)$$

$$\frac{\text{Linear}(t)}{\text{Precursor}_0} = f_{hy}\,(1 - e^{-k_T t}) \quad (2)$$

$$\frac{\text{Lariat}(t)}{\text{Precursor}_0} = f_{br}(1 - e^{-k_T t}) \quad (3)$$

$$k_{hy} = \frac{f_{hy}}{f_{hy} + f_{br}} \times k_T \quad (4)$$

$$k_{br} = \frac{f_{br}}{f_{hy} + f_{br}} \times k_T \quad (5)$$

where:

$k_T = k_{hy} + k_{br}$ = total rate of reaction

$f_{hy}$ = fraction of linear intron

$f_{br}$ = fraction of lariat intron

It is theoretically correct that $f_{hy} + f_{br} = 1$, thus simplifying

equation 4 to $k_{hy} = f_{hy} \bullet k_T$. However, it is more accurate to use the individual values for f in the denominator, as in equations 4 and 5. In so doing, one does not presume that the reaction goes to completion (a dangerous assumption since RNA degradation at long times might obscure a fraction of unreacted material). In addition, if the parallel model for splicing is correct, then the sum of $f_{hy}$ and $f_{br}$ will approach 1, thus providing an independent check on appropriate application of the kinetic model and the quality of the data (Table 5).

For reactions that followed biphasic kinetics (reactions in high concentration of monovalent ion, Figures 26A-C), equations 1-5 were adapted to account for two reactive populations of molecules:

$$\frac{\text{Precursor(t)}}{\text{Precursor}_0} = f^A \bullet e^{-k_T^A t} + f^B \bullet e^{-k_T^B t} \quad (6)$$

$$\frac{\text{Linear(t)}}{\text{Precursor}_0} = f^A_{hy} (1-+ e^{-k_T^A t}) + f^A_{hy}(1-e^{-k_T^B t}) \quad (7)$$

$$\frac{\text{Lariat(t)}}{\text{Precursor}_0} = f^A_{br} (1-+ e^{-k_T^A t}) + f^B_{br}(1-e^{-k_T^B t}) \quad (8)$$

where:

$k_T^i = k_{hy}^i + k_{br}^i$ = total reaction rate of population i (i = A, B).

$f^i$ = fraction of population$^i$ (i = A, B).

$f_{hy}^i$ = final fraction of linear product from population i (i = A,B).

$f_{br}^i$ = final fraction of lariat product from population i (i = A,B).

The experimental data in Fig. 26A-C were fit independently to equation 6-8 in order to obtain the variables $k^i_T$, $f^i_{hy}$, and $f^i_{br}$, where i = A or B. As in the simplest case presented with equations 1-3, the evolution of intronic products and the decrease in precursor were fit independently. That all four fractional values $f^A_{hy}$, $f^A_{br}$, $f^B_{hy}$, and $f^B_{br}$ add up to one and all

$k^i_T$ values for a given i are generally the same indicates that a parallel reaction model is sufficient to explain the data (Table 5). After this treatment, one can calculate the individual rates of branching or hydrolysis from each population ($k^i_{hy}$ and $k^i_{br}$, where i = A or B), using equations 9 and 10

$$k^i_{hy} = (f^A_{hy}/f^A_{hy} + f^A_{br}) \bullet k^i_T \quad (9)$$

$$k^i_{br} = (f^A_{br}/f^A_{hy} + f^A_{br}) \bullet k^i_T \quad (10)$$

**Quantitation of branched nucleotide content using P1 nuclease:** Plasmid pJD20 was transcribed as described previously, except that the amount of labeled nucleotide incorporated was increased in order to obtain a higher signal for any branched trinucleotide. Specifically, 250 uCi of a 1:1 (ethanol:water) mix of $\alpha$-$^{32}$P-GTP (Amersham) was evaporated to dryness and the components for a standard body-labeling transcription reaction were added directly into this tube. The components were: 1 $\mu$g linearized pJD20 template DNA, 2 $\mu$l of 10X transcription buffer, 0.5 mM each ATP, CTP and UTP; 0.125 mM GTP, 1 $\mu$l T7 RNA polymerase (2000 u/$\mu$l) and $H_2O$ to 20 $\mu$l. After vortexing this solution to dissolve the dried $\alpha$-$^{32}$P-GTP in the tube, the reaction was allowed to incubate for 2 hours at 37°C before loading directly on a 4% denaturing polyacrylamide gel. After electrophoresis, the full-length precursor band was visualized by autoradiography, isolated, crushed, and the RNA was eluted overnight at 4°C. The eluted RNA was recovered by ethanol precipitation.

In order to collect splicing products, 5 x $10^5$ cpm of the full length transcript was incubated under three of the standard high-salt splicing conditions described previously (either 0.5M $(NH_4)_2SO_4$, 0.5M $NH_4Cl$, or 0.5M KCl, all at 100mM $MgCl_2$, 40mM MOPS pH 7.5). Reactions were incubated at 42°C and timepoints were taken at 30 minutes (0.5M $(NH_4)_2SO_4$, and 0.5M $NH_4Cl$), 40 minutes (0.5M KCl), and 2 hours (all conditions). Precursor, lariat and putative linear products were eluted after

electrophoresis of the reactions on a 4% denaturing polyacrylamide gel. For each sample of product or precursor, the content of branched trinucleotide was analyzed by digestion with P1 nuclease (Reilly et al., 1989) in a buffer containing 25mM sodium acetate pH 5.3, 0.025 μg/ul tRNA, and 0.005 units/ul P1 nuclease (Boehringer). Reactions were incubated for 10 minutes at 37;C before adding equal volume of quench buffer (36% formamide, 0.1 X TBE, 1.8% sucrose, 0.02% xylene cyanol dye, 0.02% bromophenol blue dye, 0.02% orange G dye and 25 mM EDTA final concentrations). Samples were loaded on a denaturing 20% polyacrylamide gel. Mononucleotide pG was separated from the branched trinucleotide pA(2'-pG)-pU by electrophoresis (Chin and Pyle, 1995); gels were then dried and quantitated on a Packard Instantimager.

**References**

Alberts, B., et al. (1989) Mol. Biol. Cell. New York. Garland
Altura, R., et al. (1989) M. Nucl. Acids Res. 17:335-354
Arnberg, A. C., et al. (1980) Cell 19:313-319
Augustin, S., et al. (1990) Nature 343: 383-386
Bachl, J. and C. Schmelzer (1990) J. Mol. Biol. 212:113-125
Bachl, J. and Schmelzer, C. (1990) J. Mol. Biol. 212:113-125
Berget, S.M., et al. (1977) Proc. Natl. Acad. Sci. USA 74:3171-3175
Boguski, M.S., et al. (1979) J. Biol. Chem 255:2160-2163
Boulanger, S. C., et al. (1995) Mol. Cell. Biol. 15:4479-4488
Brody, E. and J. Abelson (1985) Science 228:963-967
Burgers, P.M.J. and Eckstein, F. (1979) Biochemistry 18:592-596
Burgess, S. M. and Guthrie, C. (1993) Cell 73:1377-1391
Cavalier-Smith, T. (1991) TIBS 7:145-148
Cech, T.R., et al. (1992) J. Biol. Chem. 267:17479-17482
Cech, T. R. (1993) Cold Spring Harbor. Cold Spring Harbor Press. 239-270
Cech, T.R., et al. (1992) J. Biol. Chem. 267:17479-17482
Cech, T.R. (1986) Cell 44:207-210
Cech, T.R. (1986) Cell 44:207-210
Chanfreau, G. and Jacquier, A. (1993) EMBO J. 2:5173:518
Chanfreau, G.and Jacquier, A. (1994) Science 266:1383-1387
Chanfreau, G. & Jacquier, A. (1994) Science 266:1383-1387
Chen, C.-y. & Sarnow, P. (1995) Science 268:415-417
Chin, K. & Pyle, A. M. (1995) RNA 1:391-406
Chow, L.T., et al. (1977) Cell 12:1-8
Costa, M. & Michel, F. (1995) EMBO 14:1276-1285
Dahm, S. C., et al. (1993) Biochemistry 32:13040-13045
Davanloo, P., et al. (1984) PNAS USA 81:2035-2039
Davies, R. W., et al. (1982) Nature 300:719-724
Dib-Hajj, S. D., et al. (1993) Nucl. Acids Res. 21:1797-1804
Donis-Keller, H., et al. (1977) Nucl. Acids REs. 4:2527-2538
Fedor, M.J. and Uhlenbeck, O.C. (1992) Biochem. 31:12042-12054
Ferat, J.-L. & Michel, F. (1993) Nature 364:358-361
Ferat, J.-L., et al. (1994) C.R. Acad. Sci. Paris. Life Sciences 317:141-148

Fersht, A. (1985) New York, W.H. Freeman, 102-109
Franzen, J.S., et al. (1993) Nucl. Acids Res. 21:627-634
Gish, G. and Eckstein, F. (1988) Science 240:1520-1522
Griffiths, A., et al. (1987) Nucl. Acids Res. 15:4145-4162
Groebe, D. R. and Uhlenbeck, O. C. (1988) Nucl. Acids Res. 16: 11725-11735
Guerrier-Takada, C., et al. (1983) Cell 35:849-857
Guthrie, C. (1991) Science 253:157-163
Halbreich, A., et al. (1980) Cell 19:321-329
Herschlag, D. and Cech, T.R. (1990a) Biochem. 29:10159-10171
Herschlag, D., et al. (1991) Biochem. 30:4844-4854
Hud, N. V., et al. (1994) Biochemistry 33:7528-7535
Jacquier, A. and Roshbash, M. (1986) Science 234:1099-1104
Jacquier, A. and Michel, F. (1987) Cell 50:17-29
Jacquier, A. and Michel, F. (1990) J. Mol. Biol. 213:437-447
Jacquier, A. and N. Jacquesson-Breuleux (1991) J. Mol. Biol. 219:415-428
Jacquier, A. (1990) TIBS 15:351-354
Jarrell, K.A., et al. (1988a) Mol. Cell. Biol. 8:2361-2366
Jarrell, K.A., et al. (1988b) J. Biol. Chem. 263:3432-3439
Jencks, W. P. (1987) New York, Dover, 569-574
Kück, U., et al. (1990) Nucl. Acids Res. 18:2691-2697
Koch, J.L., et al. (1992) Mol. Cell. Biol. 12:1950-1958
Kruger, K., et al. (1982) Cell 31:147-157
Kück, U., et al. (1990) U. Nucl. Acids Res. 18:2691-2697
Kunkel, T. A., et al. (1991) Methods Enzymol. 204:125-139
Kwakman, J.H.J.M., et al. (1989) Nucl. Acids Res. 17:4205-4216
Lambowitz, A. & Belfort, M. (1993) Ann. Rev. Biochem. 62:587-622
Lambowitz, A.M. and P.S. Perlman (1990) TIBS 15:440-444
Lambowitz, A. M. & Belfort, M. (1993) Ann. Rev. Biochem. 62:587-622
Lambowitz, A.M. (1989) Cell 56:323-326
Lamond, A.I. (1993) Nature 365:294-295
Mörl, M. & Schmelzer, C. (1990) Cell 60:629-636
Müller, M. W., et al. (1991) J. Mol. Biol. 222:145-150
Madhani, H. D. and Guthrie, C. (1992) Cell 71:803-817

Maschhoff, K.L. and Padgett, R.A. (1993) Nucl. Acids Res. 21:5456-5462

McSwiggen, J.A. and T.R. Cech (1989) Science 244:679-683

Michel, F., et al. (1989) Gene 82:5-30

Michel, F. and Dujon, B. (1983) EMBO 2:33-38

Michel, F., et al. (1982) Biochimie 64:867-881

Michels, W. J. and Pyle, A. M. (1995) Biochemistry 34:2965-2977

Michels, W. J. & Pyle, A. M. (1995) Biochemistry 34:2965-2977

Moore, M. J. & Sharp, P. A. (1992) Science 256:992-997

Moore, M.J. & Sharp, P.A. (1993) Nature 365:364-368

Moore, M.J., et al. (1993) The RNA World eds. Gesteland, R. and Atkins J., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp303-357

Moore, M.J., et al. (1993) The RNA World 1-30

Muller, M. W., et al. (1988) Nucl. Acids Res. 16:7383-7395

Muller, M. W., et al. (1991) J. Mol. Biol. 222:145-150

Pace, N. R. & Smith, D. (1990) J. Biol. Chem. 265:3587-3590

Padgett, R. A., et al. (1984) Science 225:898-903

Padgett, R. A., et al. (1994) Science 266:1685-1688

Pan, T., et al.(1993) Cold Spring Harbor, Cold Spring Harbor Press. 271-302

Peebles, C.L., et al. (1987) Cold Spring Harbor Symp. Quant. Biol. 52:223-232

Peebles, C.L., et al. (1986) Cell 44:213-223

Peebles, C.L., et al. (1993) J. Biol. Chem 268:11929-11938

Perlman, P.S. and R.A. Butow (1989) Science 246:1106-1109

Perlman, P. S., et al. (1986) New York, Plenum press. 39-55

Podar, M., et al. (1995b) RNA , in press.

Podar, M., et al. (1995a) Mol. Cell. Biol. 15:4466-4478

Potter, V.L., et al. (1983) Biochem 22:1369-1376

Pyle, A.M., et al. (1992) Nature 358:123-128

Pyle, A.M., et al. (1990) Proc. Natl. Acad. Sci. USA 87:8187-8191

Pyle, A.M. and T.R. Cech (1991) Nature 350:628-631

Pyle, A.M. and Green, J.B. (1994) Biochem. 33:2716-2725

Rajogopal, J., et al. (1989) Science 244:692-694

Reich, C., et al. (1988) Science 239:117-232

Reilly, J. D., et al. (1989) Methods Enzymol. 180:177-191
Reilly, J. D., et al. (1989) Methods Enzymol. 180:177-191
Saldhana, R., et al. (1993) FASEB 7:15-24
Scaringe, S.A., et al. (1990) Nucl. Acids Res. 18:5433-5441
Schatz, D., et al. (1991) Proc. Natl. Acad Sci. USA 88:6132-6136
Schmelzer, C., et al. (1982) Nucl. Acids Res. 10:6797-6808
Schmelzer, C. and Schweyen, R. J. (1986) Cell 46:557-565
Schmelzer, C. & Müller, M. W. (1987) Cell 51:753-762
Schmidt, U., et al. (1987) Current Genetics 12:291-295
Sharp, P.A. (1985) Cell 42:397-400
Suchy, M. and Schmelzer, C. (1991) J. Mol. Biol. 222:179-187
Sulston, J. R., et al. PNAS USA 59:726-733
Tinoco, I., et al. (1978) Englewood Cliffs, N.J., Prentice Hall
Uhlenbeck, O. C. (1995) RNA 1:4-6
Umesono, M.F. and Ozeki, H. (1989) Gene 82:5-30
van der Veen, R., et al. (1987) EMBO Journal 6:3827-3831
van der Veen, R., et al. (1986) Cell 44:225-234
Van Dyck, E., et al. (1995) Mol. & General Gen. 246:426-436
Walsh, C. (1979) San Francisco, W.H. Freeman.
Walstrum, S. A. & Uhlenbeck, O. C. (1990) Biochemistry 29:10573-10576
Warrant, R. W. & Kim, S. H. (1978) Nature 271:130-135
Wiesenberger, G., et al. (1992) J. Biol. Chem. 267:6963-6969
Winkler, M. & Kuck, U. (1991) Current Genetics 20:495-502
Wissinger, B., et al. (1992) Trends in Genetics 8:322-328
Woodson, S. A. and Cech, T. R. (1989) Cell 57:335-345
Zaug, A.J., et al. (1986) Science 324:429-433
Zaug, A.J. and Cech, T.R. (1986) Science 231:470-475
Zaug, A. J., et al. (1994) Biochemistry 33:14935-14949

WHAT IS CLAIMED IS:

1. A composition for catalyzed oligonucleotide cleavage comprising a synthetic non-naturally occurring oligonucleotide compound which compound comprises nucleotides whose sequence defines a conserved group II intron catalytic region and nucleotides whose sequence is capable of hybridizing with a predetermined oligonucleotide target sequence to be cleaved, such target sequence not being present within the compound, and an appropriate oligonucleotide co-factor.

2. The composition of claim 1, wherein the conserved group II intron catalytic region is a group II intron domain I catalytic region.

3. The composition of claim 2, wherein the conserved group II intron domain I catalytic region further comprises a conserved portion of a group II intron domain II, a group II intron domain III, a group II intron domain IV, a group II intron domain V, or a group II intron domain VI in cis or trans.

4. The composition of claim 1, wherein the nucleotides whose sequence is capable of hybridizing with a predetermined oligonucleotide target sequence to be cleaved comprises two hybridizing regions each region having 2 to 12 nucleotides capable of hybridizing with the oligonucleotide target to be cleaved.

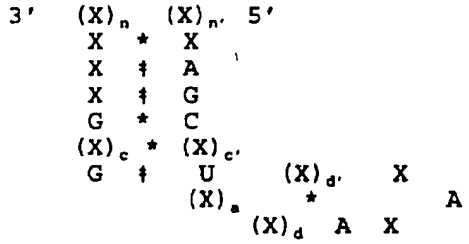
5. The composition of claim 4, wherein each hybridizing regions has 6 or 7 nucleotides.

6. The composition of claim 1, wherein the cleavage is due to transesterification.

7. The composition of claim 1, wherein the synthetic non-

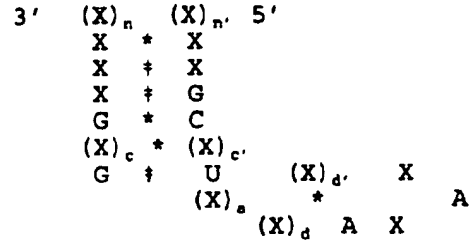
naturally occurring oligonucleotide compound comprises between about 80 and about 1000 nucleotides.

8. The composition of claim 1 wherein the appropriate co-factor is an oligonucleotide compound from a conserved portion of a group II intron domain V having the formula:




wherein each X represents a ribonucleotide which may be the same or different; wherein each of $(X)_n$, $(X)_{n'}$, $(X)_c$, $(X)_{c'}$, $(X)_{d'}$, $(X)_a$ and $(X)_d$ represents an oligonucleotide; wherein n, n', c, c', d, d', and a each represents an integer which defines the number of nucleotides in the oligonucleotide with the provisos that n and n' are greater than or equal to 1; a represents an integer which is greater than or equal to 1; d and d' represent an integer which is greater than or equal to 5; c and c' represents an integer which is greater than or equal to 4; wherein each * represents base pairing between the nucleotide located on either side thereof; wherein each ǂ may or may not represent base pairing between the nucleotide located on either side thereof; and wherein each solid line represents a chemical linkage providing covalent bonds between the nucleotide located on either side thereof.

9. The composition of claim 1 wherein the appropriate co-factor is an oligonucleotide compound from a conserved portion of a group II intron domain V having the formula:



wherein each X represents a ribonucleotide which may be the same or different; wherein each of $(X)_n$, $(X)_{n'}$, $(X)_c$, $(X)_{c'}$, $(X)_{d'}$, $(X)_a$ and $(X)_d$ represents an oligonucleotide; wherein n, n', c, c', d, d', and a each represents an integer which defines the number of nucleotides in the oligonucleotide with the provisos that n and n' are greater than or equal to 1; a represents an integer which is greater than or equal to 1; d and d' represent an integer which is greater than or equal to 5; c and c' represents an integer which is greater than or equal to 4; wherein each * represents base pairing between the nucleotide located on either side thereof; wherein each ‡ may or may not represent base pairing between the nucleotide located on either side thereof; and wherein each solid line represents a chemical linkage providing covalent bonds between the nucleotide located on either side thereof.

10. The composition of claim 1 further comprising a divalent cation.

11. The divalent cation of claim 10, wherein the divalent cation is $Mg^{2+}$ or $Mn^{2+}$.

12. The composition of claim 1, wherein the predetermined oligonucleotide target sequence to be cleaved is mRNA.

13. The composition of claim 1, wherein the predetermined

oligonucleotide target sequence to be cleaved is DNA.

14. The composition of claim 13, wherein the mRNA encodes a growth factor.

15. The composition of claim 14, wherein the growth factor is an angiogenic factor, a basic fibroblast growth factor, a colony-stimulating factor 1, cystic fibrosis transmembrane conductance regulator, an epidermal growth factor, an erythropoietin, a fibroblast growth factor, a G-protein, a granulocyte-macrophage colony stimulating factor, a growth hormone, IL-1, IL-1R, IL-2, IL-2R, IL-4, IL-6, an insulin-like growth factor, an insulin-like growth factor 1, an interferon, an interleukin, a keratinocyte growth factor, luteinizing hormone receptor, MDR-1, a nerve growth factor, a platelet derived growth factor, a scatter factor, a transforming growth factor $\alpha$, a transforming growth factor $\beta$, a transforming growth factor, or a tumor necrosis factor.

16. The composition of claim 12, wherein the mRNA encodes an oncogene or a tumor suppressor gene.

17. The mRNA of claim 16, wherein the oncogene or tumor suppressor gene is *bcl*-2, *bcr-abl*, *bek*, BPV, c-*abl*, c-*fes*, c-*fms*, c-*fos*, c-H-*ras*, c-*kit*, c-*myb*, c-*myc*, c-*mos*, c-*sea*, *cerbB*, DCC, erbA, *erbB*-2, *ets*, *fig*, FSFV gp55, Ha-*ras*, HIV *tat*, HTLV-1 *tat*, JCV early, *jun*, L-*myc*, *lck*, LPV early, *met*, N-*myc*, NF-1, N-*ras*, *neu*, p53, Py mTag, *pim*-1, *ras*, RB, *rel*, retinoblastoma-1, SV-40 Tag, TGF-$\alpha$, TGF-$\beta$, *trk*, *trkB*, *v-abl*, v-H-*ras*, *v-jun*, or WT-1.

18. The composition of claim 17, wherein the mRNA is an mRNA associated with a chromosomal translocation.

19. The composition of claim 17, wherein the oncogene or tumor suppressor gene has one or more point mutations.

20. The composition of claim 12, wherein the mRNA is an mRNA whose overproduction is associated with a disease or condition.

21. The composition of claim 12, wherein the mRNA is a mammalian mRNA.

22. The composition of claim 12, wherein the mRNA is a yeast mRNA.

23. The composition of claim 12, wherein the mRNA is a bacterial mRNA.

24. The composition of claim 12, wherein the mRNA is a viral mRNA.

25. The composition of claim 24, wherein the viral mRNA is associated with a cytomegalovirus, an Epstein-Barr virus, a hepatitis B virus, a hepatitis C virus, a herpes simplex virus, a herpesvirus, a HIV-1 virus, an immunodeficiency virus, an influenza virus, a papillomavirus, a picornavirus, a polio virus or a T-cell leukemia virus.

26. The composition of claim 12, wherein the mRNA is a plant mRNA.

27. The composition of claim 26, wherein the plant mRNA is an mRNA associated with alfalfa, apples, asparagus, bananas, broccoli, carrots, celery, chicory, coffee, cabbage, mustard, corn, cottonseed, squash, cucumber, cantaloupe, grapes, lettuce, palm, potato, rapeseed, raspberry, soybean, sunflower, strawberry, tomato, or wheat.

28. An oligonucleotide transfer vector containing a nucleotide sequence which on transcription gives rise to the synthetic non-naturally occurring compound of claim 1.

29. The transfer vector of claim 28, wherein the transfer vector is a bacterial plasmid, a bacteriophage DNA, a cosmid, or an eukaryotic viral DNA.

30. A host cell transformed by the transfer vector of claim 28.

31. The host cell of claim 30, wherein the host cell is a prokaryotic host cell or an eukaryotic host cell.

32. The prokaryotic host cell of claim 31, wherein the prokaryotic host cell is an E. coli host cell.

33. The eukaryotic host cell of claim 31, wherein the eukaryotic host cell is a monkey COS host cell, a Chinese hamster ovary host cell, a mammalian host cell or a plant host cell.

34. A method for cleaving a target oligonucleotide comprising the steps of:

a) composition for catalyzed oligonucleotide cleavage comprising a synthetic non-naturally occurring oligonucleotide compound which compound comprises nucleotides whose sequence defines a conserved group II intron catalytic region and nucleotides whose sequence is capable of hybridizing with a predetermined oligonucleotide target sequence to be cleaved, such target sequence not being present within the compound, and an appropriate oligonucleotide co-factor and

b) contacting the synthetic non-naturally occurring oligonucleotide compound with a target oligonucleotide to cause cleavage of the target oligonucleotide.

35. The method for cleaving an oligonucleotide of claim 34, wherein the conserved group II intron catalytic region further comprises a group II intron domain II, a group II intron domain III, a group II intron domain IV, or a group II intron domain VI.

36. The method for cleaving an oligonucleotide of claim 34, wherein the nucleotides whose sequence is capable of hybridizing with a predetermined oligonucleotide target sequence to be cleaved comprises two hybridizing regions each region having 2 to 10 nucleotides.

37. The two hybridizing regions of claim 36 having 6 and 7 nucleotides respectively.

38. The method for cleaving an oligonucleotide of claim 34 wherein the composition further comprises a divalent cation.

39. The divalent cation of claim 38, wherein the divalent cation is $Mg^{2+}$ or $Mn^{2+}$.

40. The method for cleaving an oligonucleotide of claim 34, wherein the predetermined oligonucleotide target sequence to be cleaved is mRNA.

41. A transgenic, non-human vertebrate animal, having one or more cells bearing a DNA sequence encoding the composition of claim 1.

42. The transgenic animal of claim 41, wherein the animal is a mammal.

43. The transgenic mammal of claim 42, wherein the mammal is selected from the group consisting of cows, goats, sheep, pigs, horses, dogs, cats, and rodents.

44. The transgenic animal of claim 41, wherein the animal is a fish, a chicken, or a turkey.

45. The transgenic animal of claim 41, wherein the target sequence is a viral target sequence.

46. The viral target sequence of claim 45, wherein the viral target sequence is a target sequence from a picornavirus, an immunodeficiency virus, a hepatitis B virus, a papillomavirus, an Epstein-Barr virus, a T-cell leukemia virus, a hepatitus C virus, a cytomegalovirus, an influenza virus, a herpes simplex virus, or a herpesvirus.

47. The transgenic animal of claim 41, wherein the target sequence is a target sequence from a mRNA encoding a growth factor.

48. The transgenic animal of claim 47, wherein the growth factor is an insulin-like growth factor, insulin-like growth factor 1, a transforming growth factor, transforming growth factor $\beta$, transforming growth factor $\alpha$, an interferon, a tumor necrosis factor, a platelet derived growth factor, a nerve growth factor, an epidermal growth factor, a fibroblast growth factor, a basic fibroblast growth factor, colony-stimulating factor 1, an interleukin, a granulocyte-macrophage colony stimulating factor, erythropoietin, a HGF, a keratinocyte growth factor, a scatter factor, or an angiogenic factor.

49. The transgenic mammal of claim 41, wherein the target sequence is a target sequence from a mRNA encoding an oncogene or a tumor suppressor gene.

50. The mRNA of claim 49, wherein the oncogene or tumor suppressor gene is *bcl*-2, *bcr-abl*, *bek*, BPV, c-*abl*, c-*fes*, c-*fms*, c-*fos*, c-H-*ras*, c-*kit*, c-*myb*, c-*myc*, c-*mos*, c-*sea*, *cerbB*, DCC, erbA, *erbB*-2, *ets*, *fig*, FSFV gp55, Ha-*ras*,

HIV *tat*, HTLV-1 *tat*, JCV early, *jun*, L-*myc*, *lck*, LPV early, *met*, N-*myc*, NF-1, N-*ras*, *neu*, p53, Py mTag, *pim*-1, *ras*, RB, *rel*, retinoblastoma-1, SV-40 Tag, TGF-$\alpha$, TGF-$\beta$, *trk*, *trkB*, *v-abl*, v-H-*ras*, *v-jun*, or WT-1.

51. A method of treatment which comprises:
    a) recovering a subject's appropriate cells;
    b) transforming the cells with a DNA sequence encoding the composition of claim 1; and
    c) introducing the resulting transformed cells into the subject so as to treat the subject.

52. The method of claim 51, wherein the target sequence is a viral target sequence.

53. The viral target sequence of claim 52, wherein the viral target sequence is a target sequence from a picornavirus, an immunodeficiency virus, a hepatitis B virus, a papillomavirus, an Epstein-Barr virus, a T-cell leukemia virus, a hepatitus C virus, a cytomegalovirus, an influenza virus, a herpes simplex virus, or a herpesvirus.

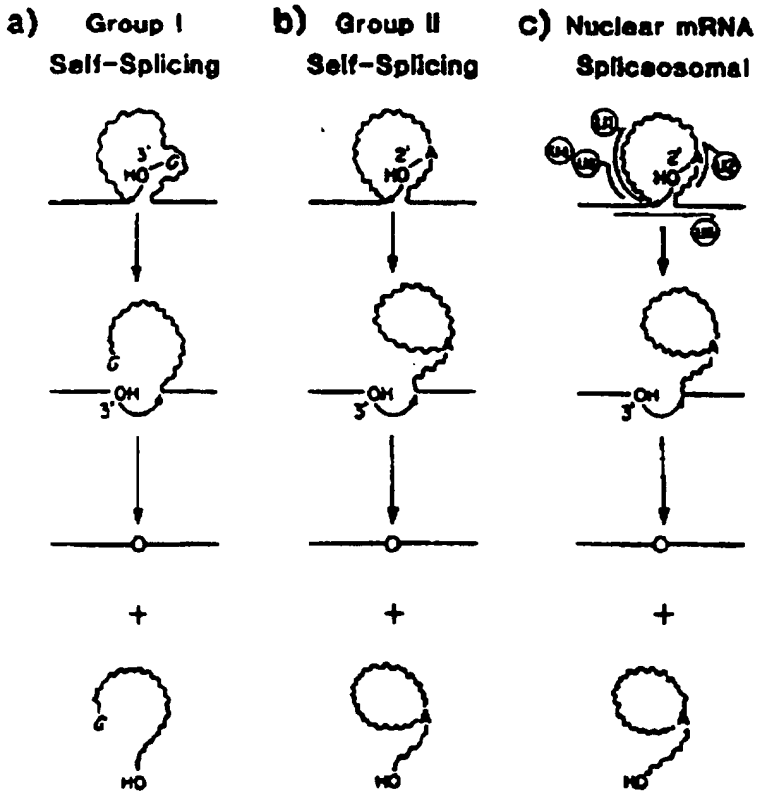
54. The method of claim 52, wherein the target sequence is a mRNA encoding a growth factor.

55. The method of claim 54, wherein the growth factor is a target sequence from an insulin-like growth factor, insulin-like growth factor 1, a transforming growth factor, transforming growth factor $\beta$, transforming growth factor $\alpha$, an interferon, a tumor necrosis factor, a platelet derived growth factor, a nerve growth factor, an epidermal growth factor, a fibroblast growth factor, a basic fibroblast growth factor, colony-stimulating factor 1, an interleukin, a granulocyte-macrophage colony stimulating factor, erythropoietin, a HGF, a keratinocyte growth factor, a scatter factor, or an angiogenic factor.

56. The method of claim 52, wherein the target sequence is a target sequence from a mRNA encoding an oncogene or a tumor suppressor gene.

57. The mRNA of claim 56, wherein the oncogene or tumor suppressor gene is *bcl*-2, *bcr-abl*, *bek*, BPV, c-*abl*, c-*fes*, c-*fms*, c-*fos*, c-H-*ras*, c-*kit*, c-*myb*, c-*myc*, c-*mos*, c-*sea*, *cerbB*, DCC, erbA, *erbB*-2, *ets*, *fig*, FSFV gp55, Ha-*ras*, HIV *tat*, HTLV-1 *tat*, JCV early, *jun*, L-*myc*, *lck*, LPV early, *met*, N-*myc*, NF-1, N-*ras*, *neu*, p53, Py mTag, *pim*-1, *ras*, RB, *rel*, retinoblastoma-1, SV-40 Tag, TGF-$\alpha$, TGF-$\beta$, *trk*, *trkB*, *v-abl*, *v*-H-*ras*, *v-jun*, or WT-1.

FIGURE 1


a) Group I Self-Splicing
b) Group II Self-Splicing
c) Nuclear mRNA Spliceosomal
3'
HO
G
OH
3'
+
G
HO
2'
HO
A
OH
3'
+
A
HO
U1
U4
U6
U5
2'
HO
A
U2
U6
OH
3'
+
A
HO

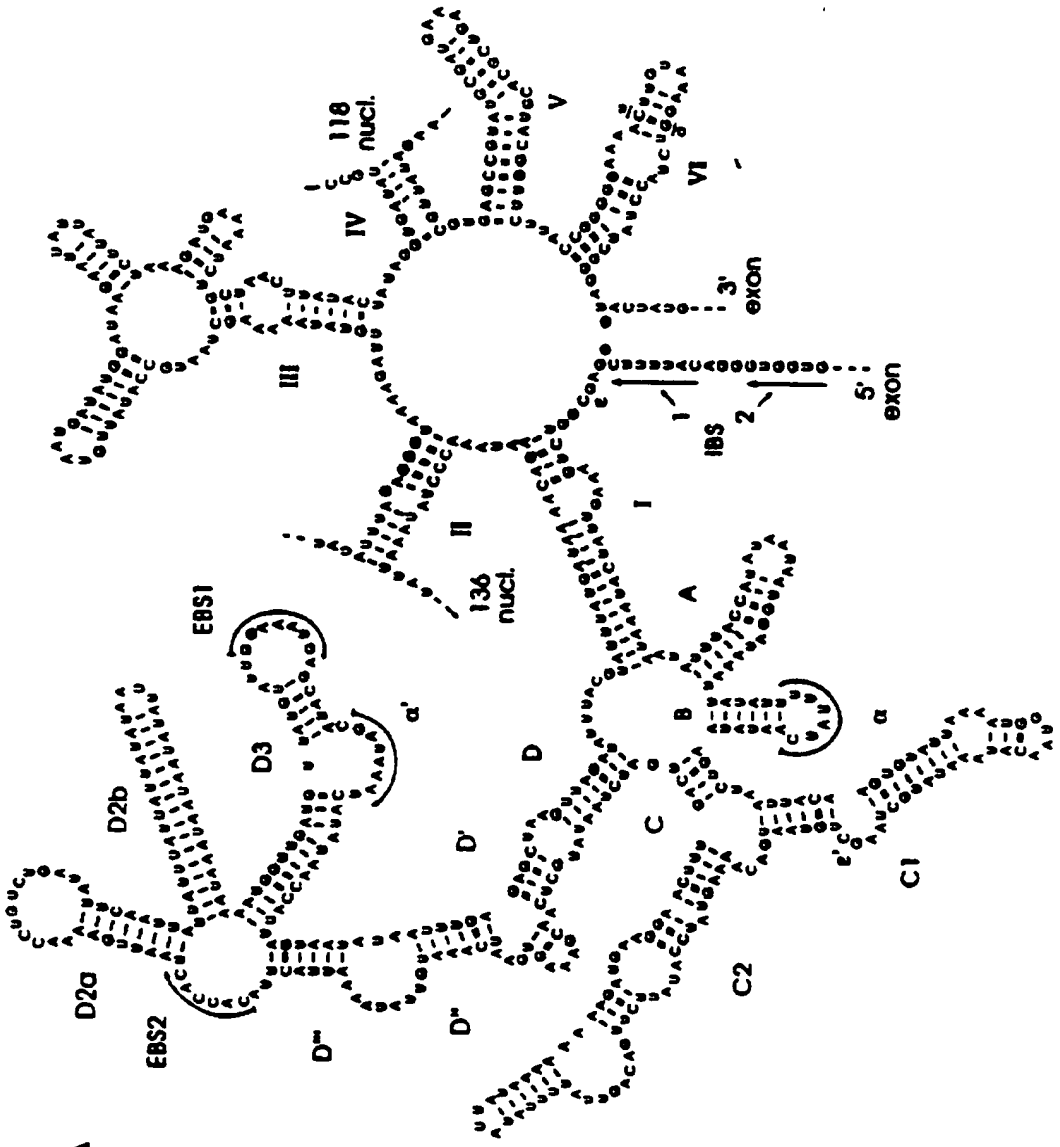

118 nucl.
IV
V
VI
3' exon
5' exon
III
1
2
IBS
I
II
136 nucl.
EBS1
A
B
α
α'
D3
D2b
D
D'
C
C1
C2
D2a
EBS2
D'''
D''

FIGURE 2A

## FIGURE 2B

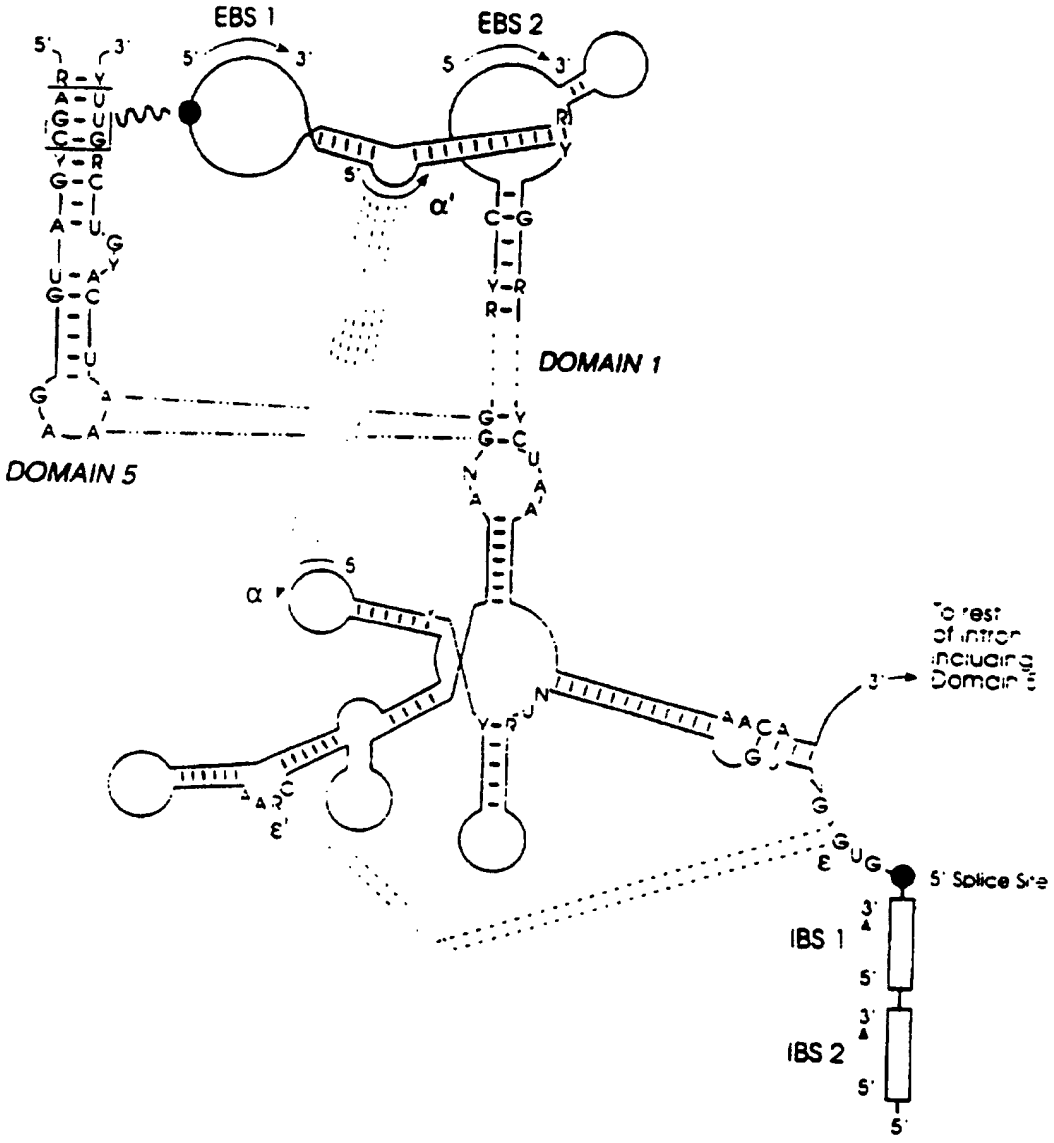


FIGURE 3

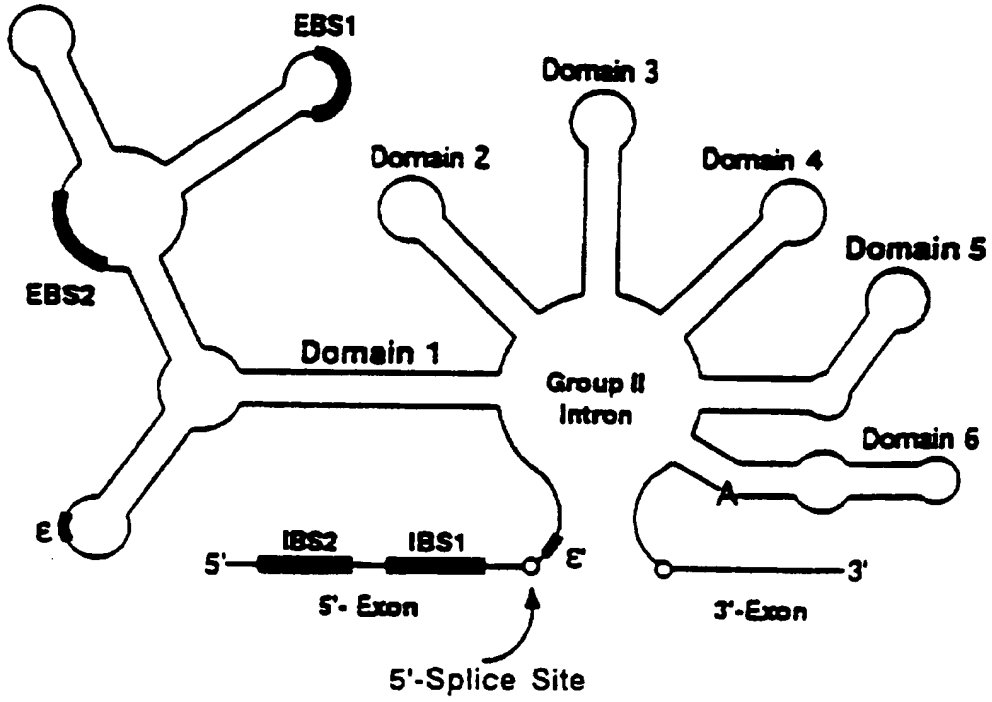

EBS1
Domain 3
Domain 2
Domain 4
Domain 5
EBS2
Domain 1
Group II
Intron
Domain 6
A
ε
IBS2
IBS1
ε'
5'
3'
5'- Exon
3'-Exon
5'-Splice Site

FIGURE 4A

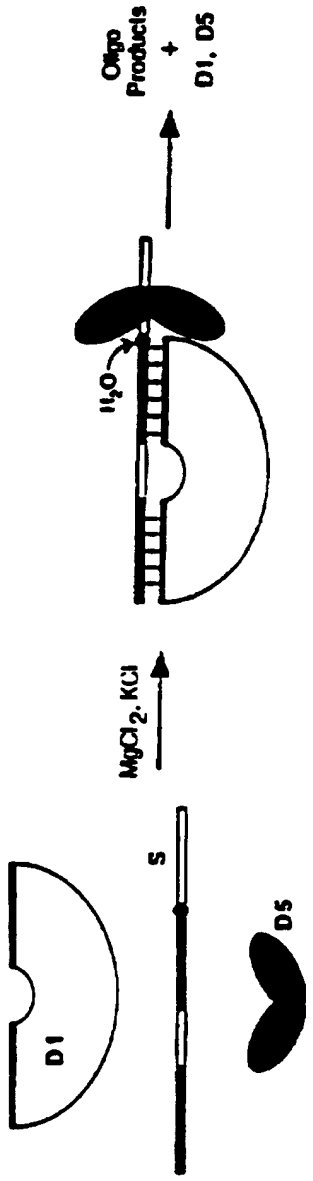
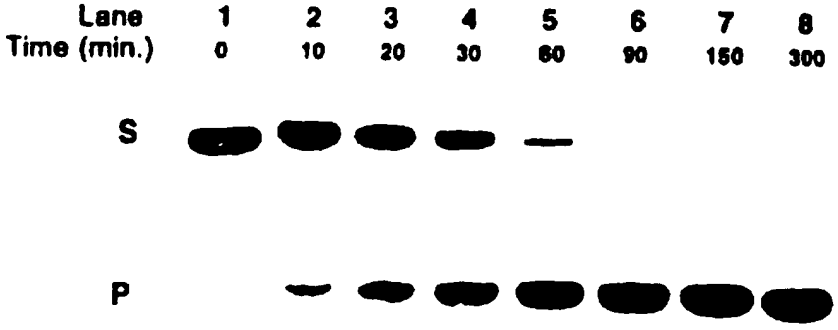


FIGURE 4B



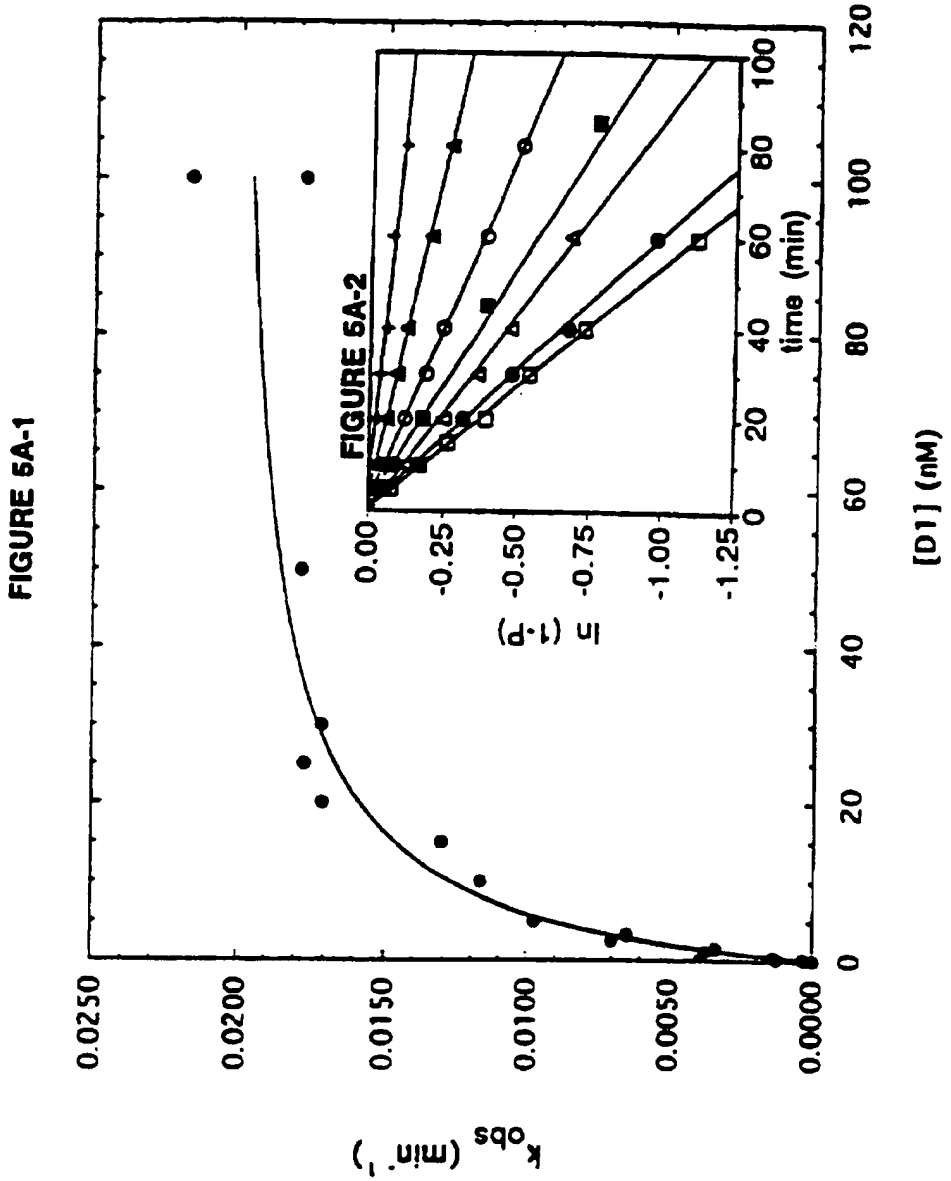


FIGURE 5A-1
FIGURE 5A-2
$k_{obs}$ (min$^{-1}$)
[D1] (nM)
time (min)
$\bar{n}$ (1-$\bar{P}$)
0.0250
0.0200
0.0150
0.0100
0.0050
0.0000
0
20
40
60
80
100
120
0.00
-0.25
-0.50
-0.75
-1.00
-1.25

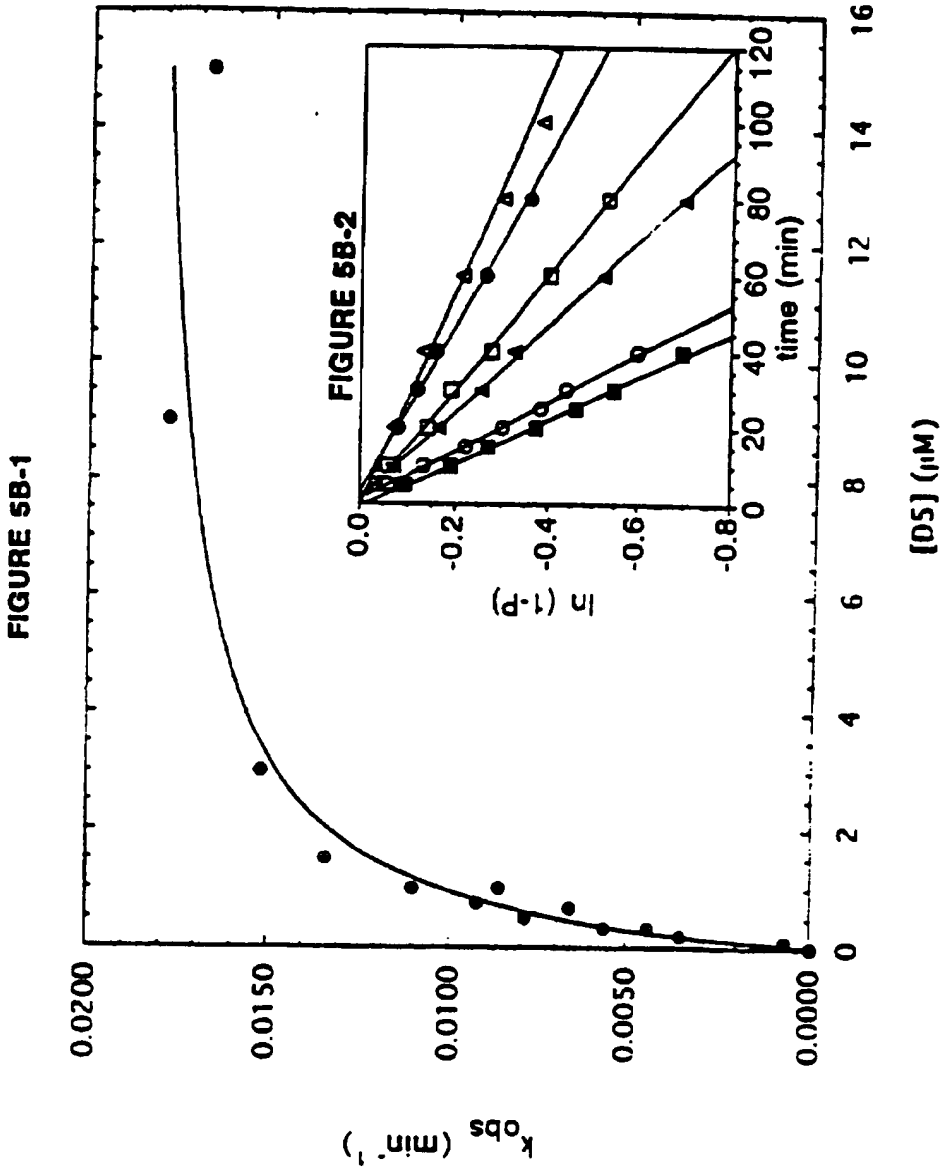


FIGURE 5B-1
FIGURE 5B-2
$k_{obs}$ ($min^{-1}$)
0.0200
0.0150
0.0100
0.0050
0.0000
0
2
4
6
8
10
12
14
16
[DS] (μM)
ln (1-P)
0.0
-0.2
-0.4
-0.6
-0.8
0
20
40
60
80
100
120
time (min)

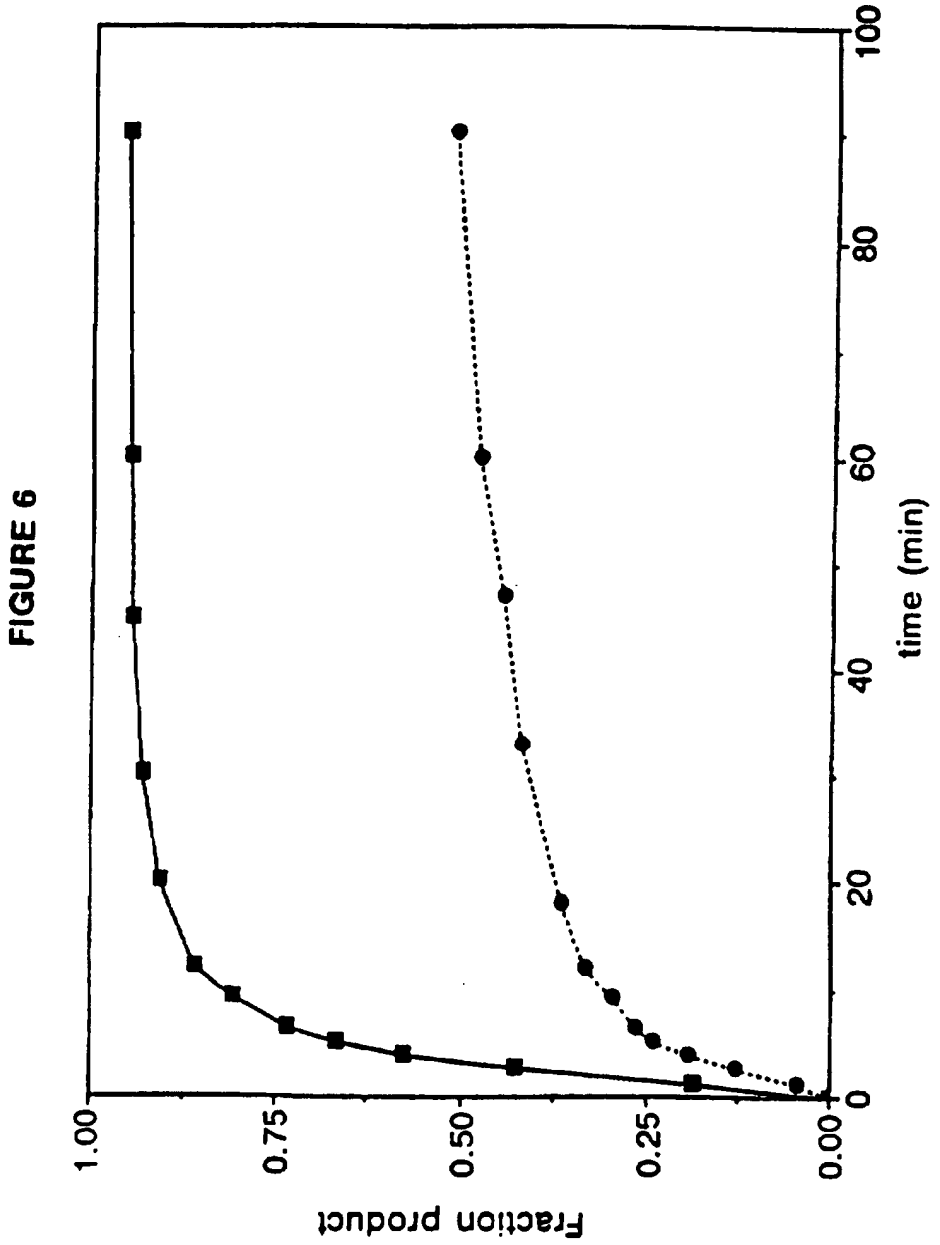



FIGURE 6

FIGURE 7

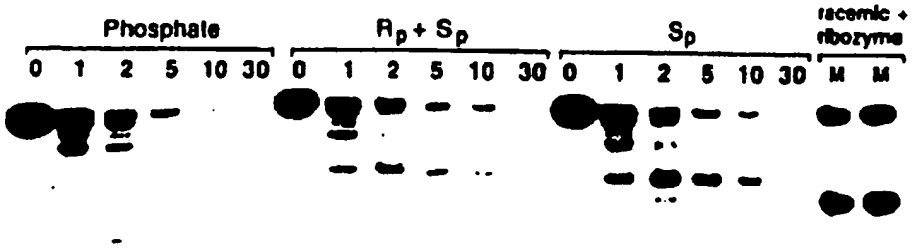

Phosphate
$R_p + S_p$
$S_p$
racemic + ribozyme
0 1 2 5 10 30 0 1 2 5 10 30 0 1 2 5 10 30 M M

FIGURE 8A

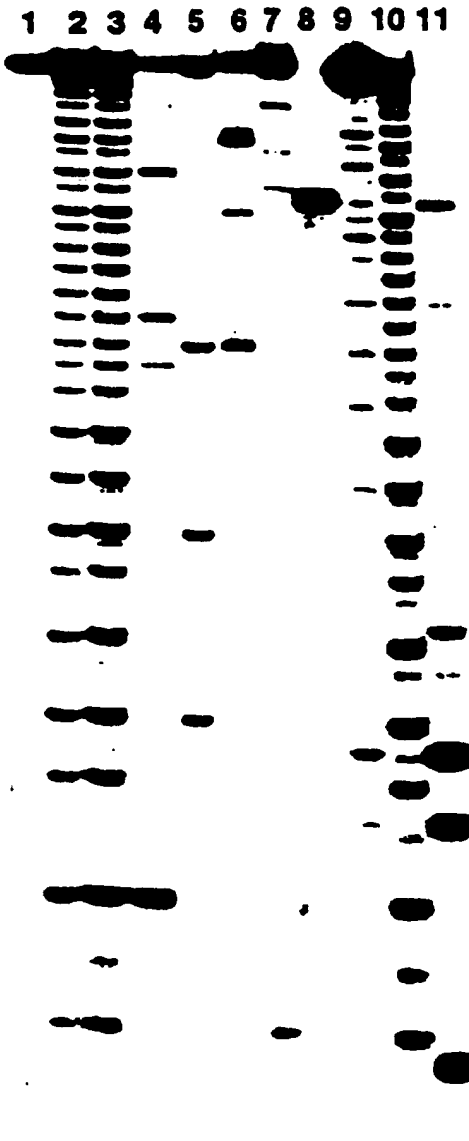



FIGURE 8B

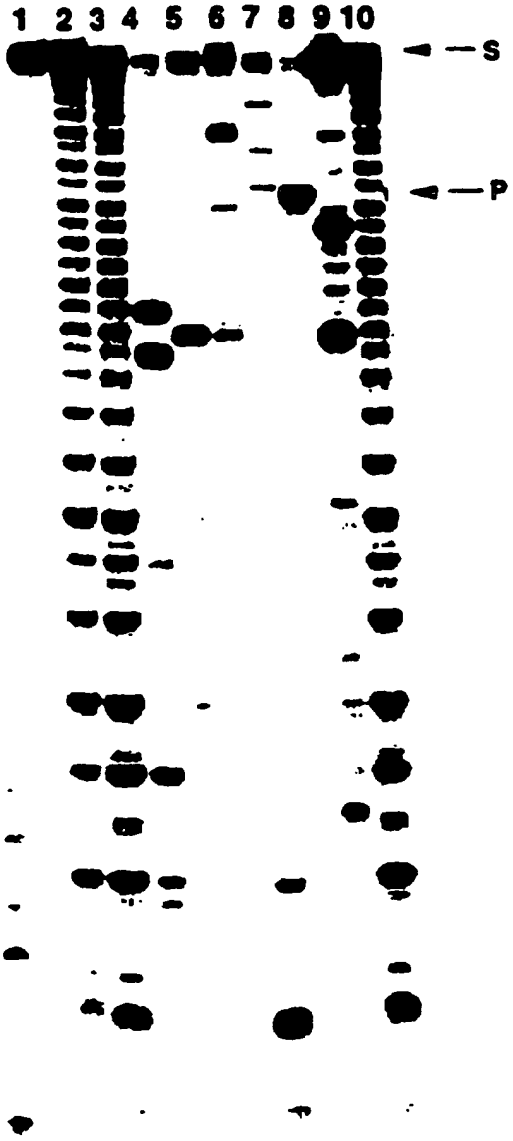


FIGURE 9A

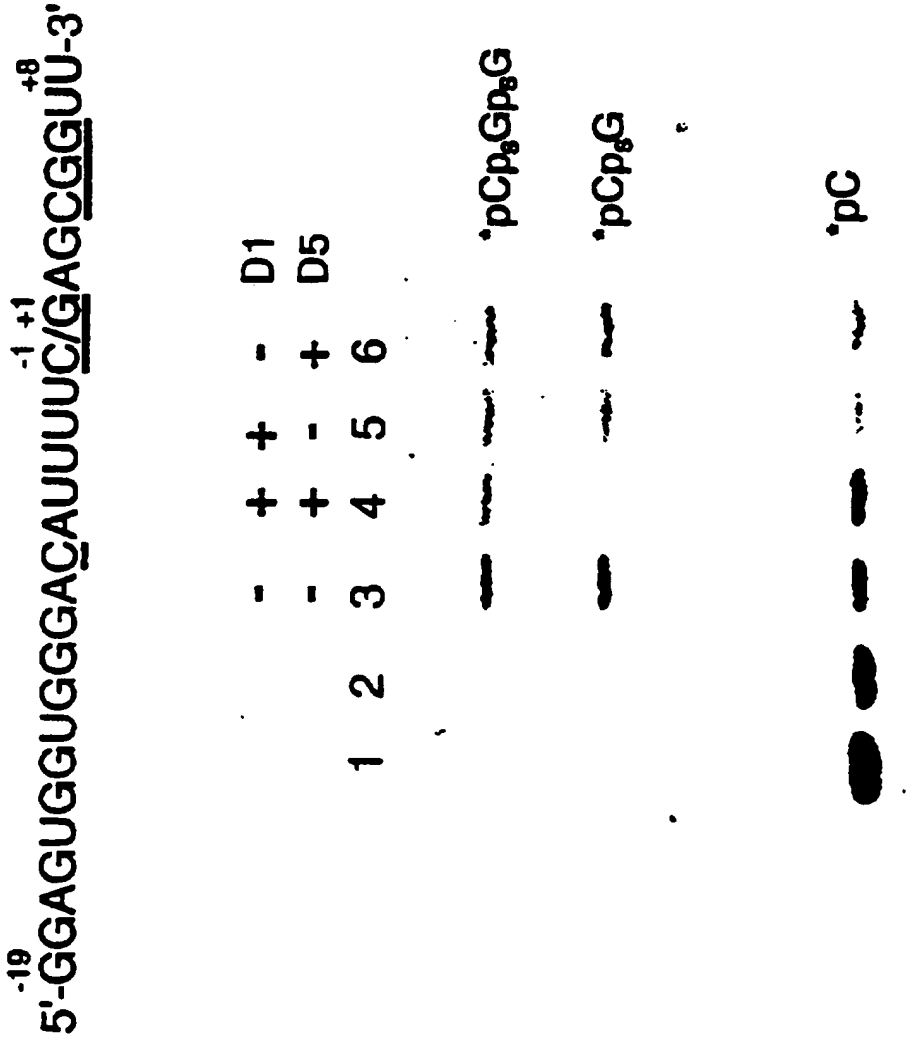

-19
5'-GGAGUGGUGGGACAUUUUC/GAGCGGUU-3'
-1 +1
+8
1 2 3 4 5 6
- + - + - + D1
- + + - + - D5
*pCp8Gp8G
*pCp8G
*pC

FIGURE 9B

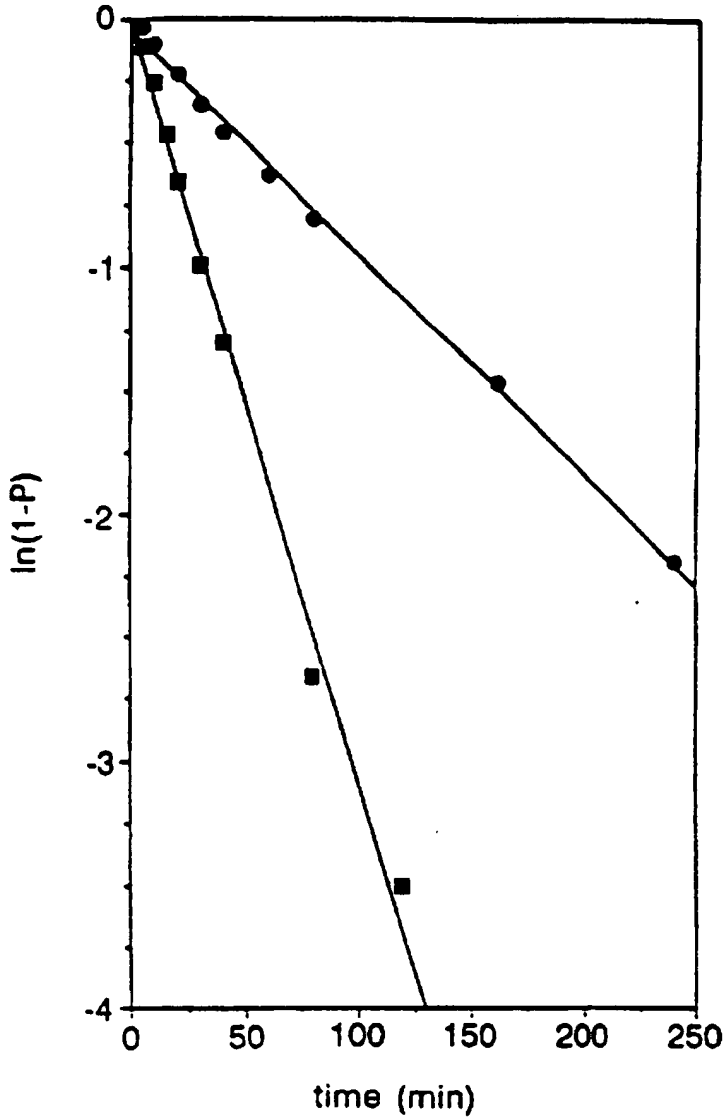


FIGURE 10A

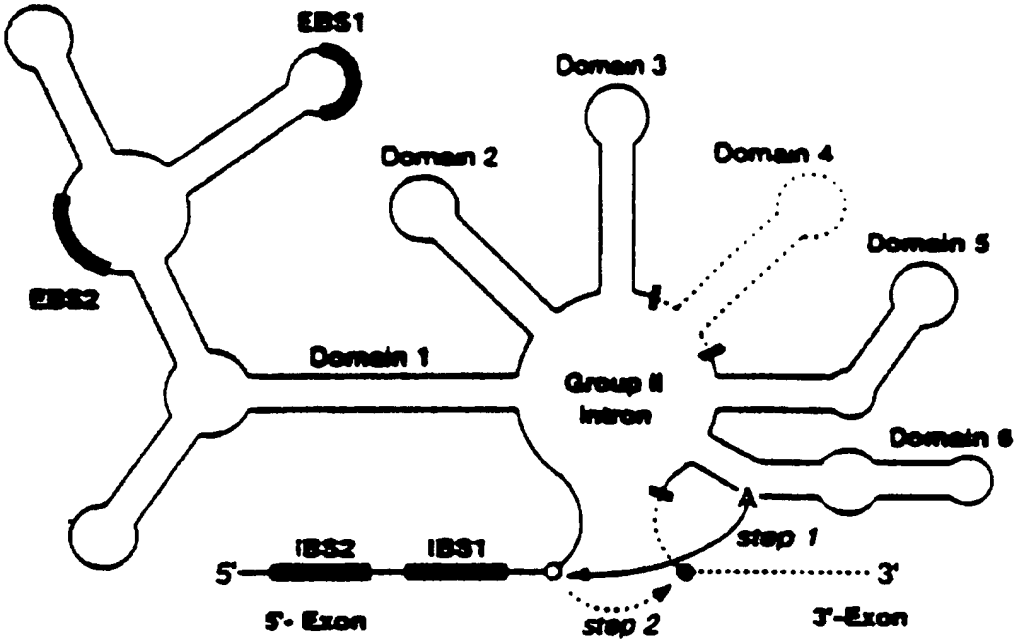



FIGURE 10B

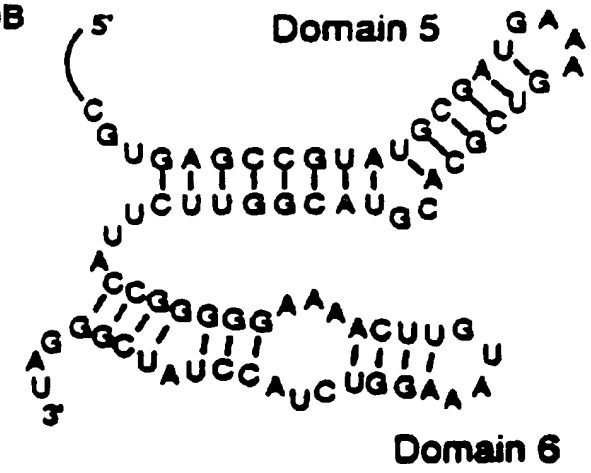


FIGURE 11A

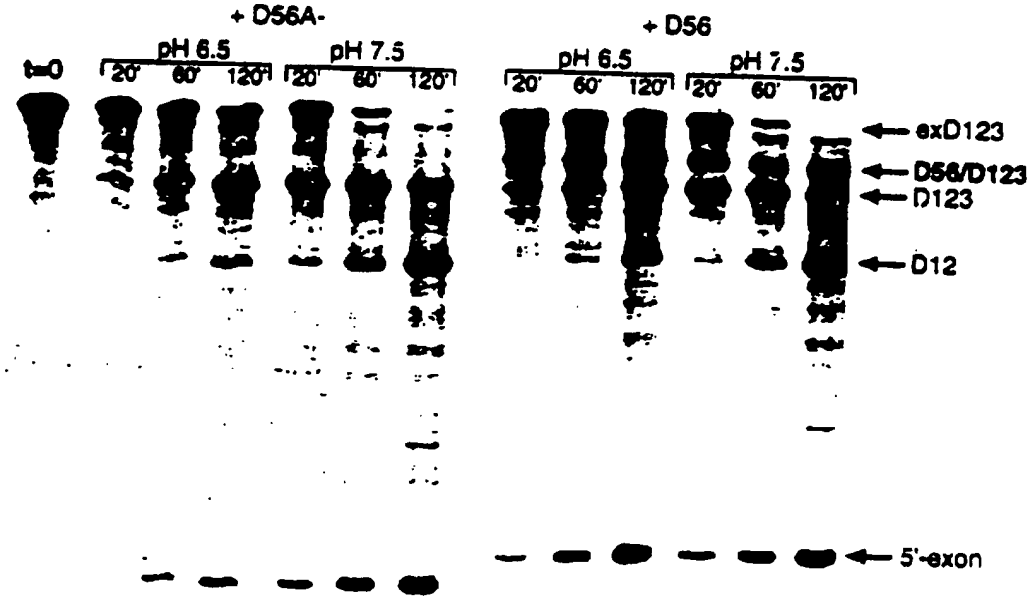


FIGURE 11C

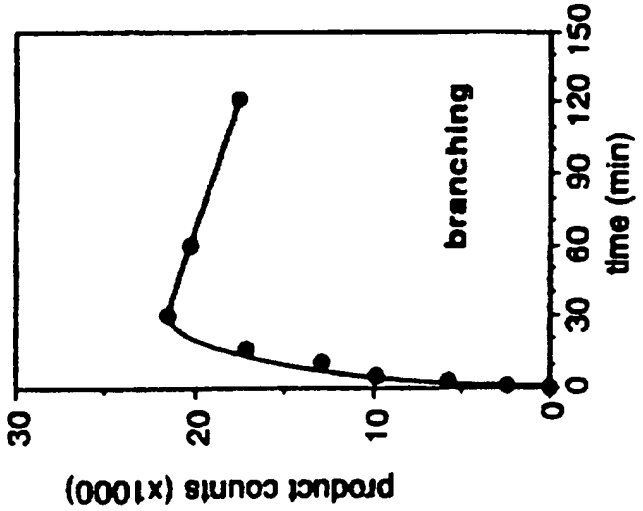
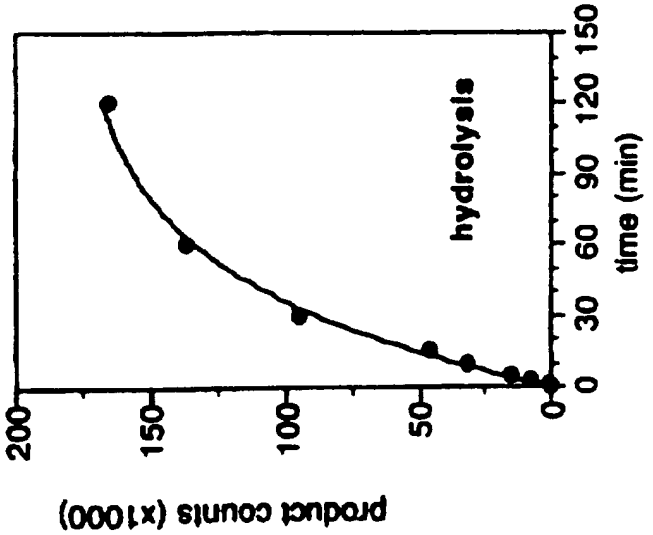



FIGURE 11B



## FIGURE 12A

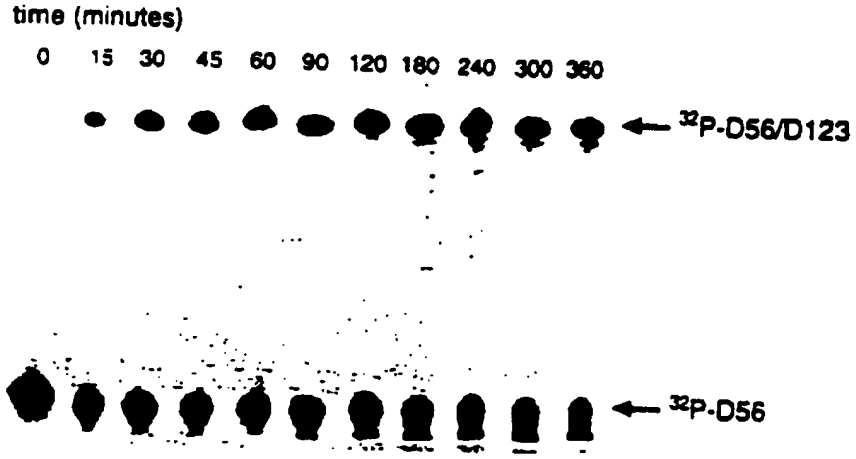

time (minutes)
0 15 30 45 60 90 120 180 240 300 360
32P-D56/D123
32P-D56

FIGURE 12B

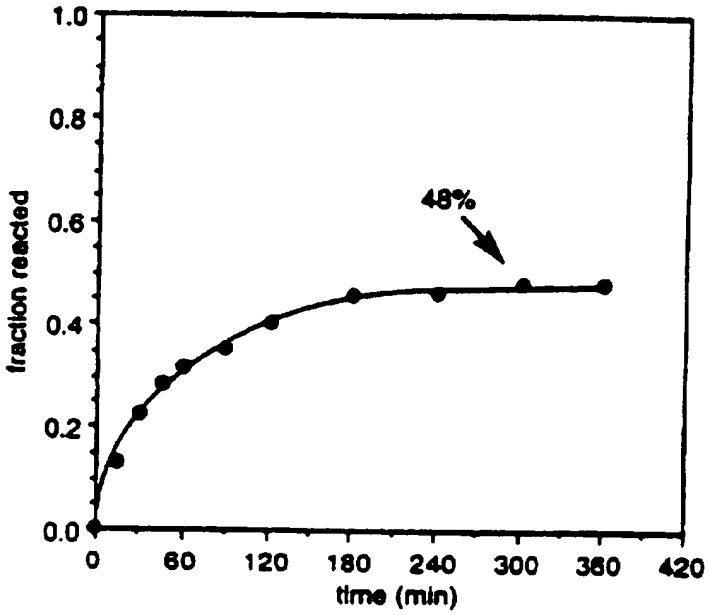

1.0
0.8
0.6
0.4
0.2
0.0
fraction reacted
48%
0
60
120
180
240
300
360
420
time (min)

FIGURE 12C

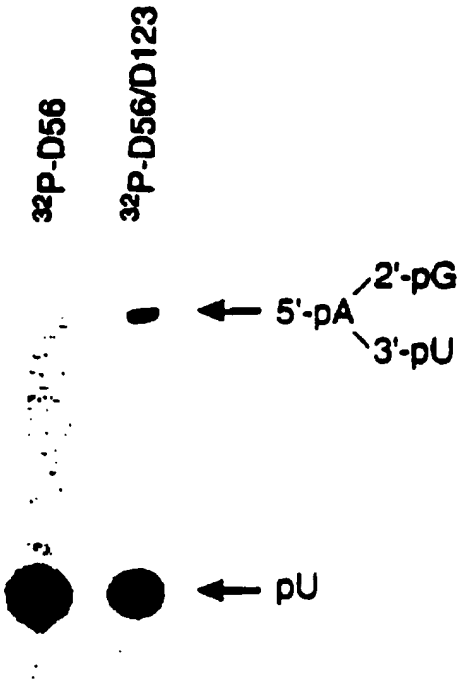


FIGURE 13A

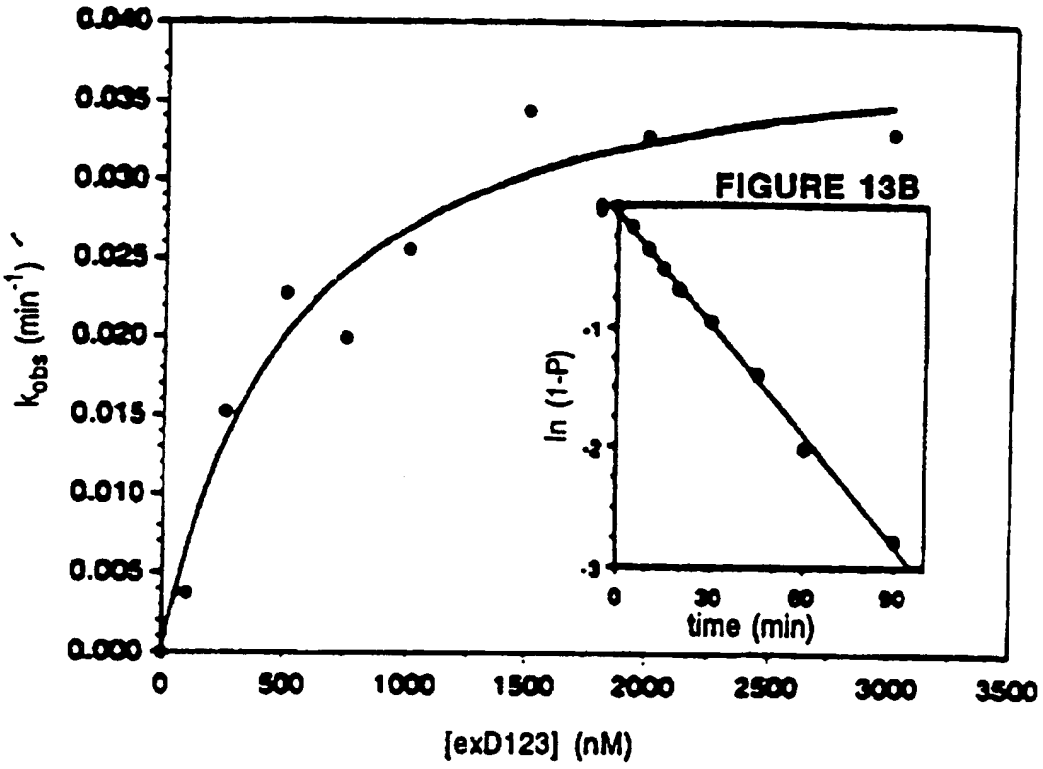
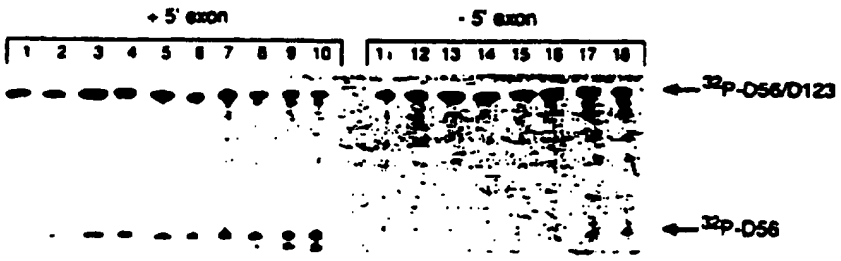

0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
$k_{obs}$ ($min^{-1}$)
0
500
1000
1500
2000
2500
3000
3500
[exD123] (nM)
FIGURE 13B
ln (1-P)
-1
-2
-3
0
30
60
90
time (min)

FIGURE 14A


+ 5' exon
- 5' exon
1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18
32P-D56/D123
32P-D56

## FIGURE 14B

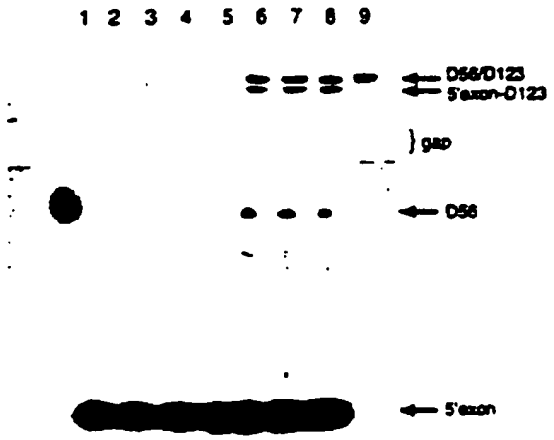
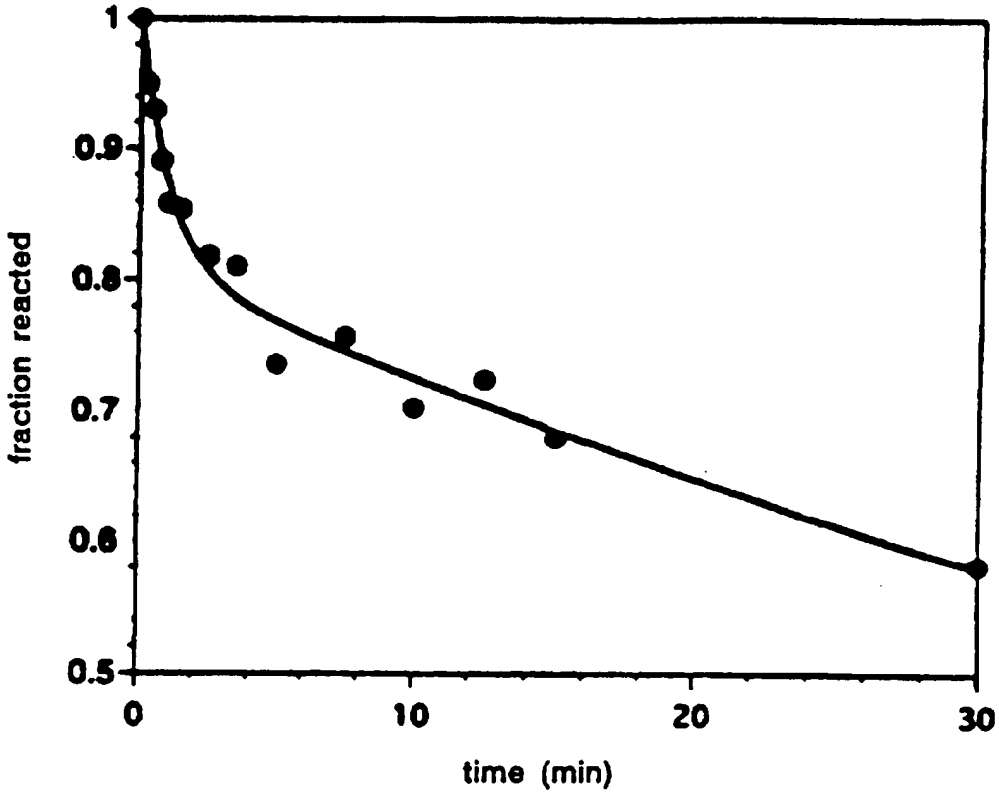


FIGURE 14C



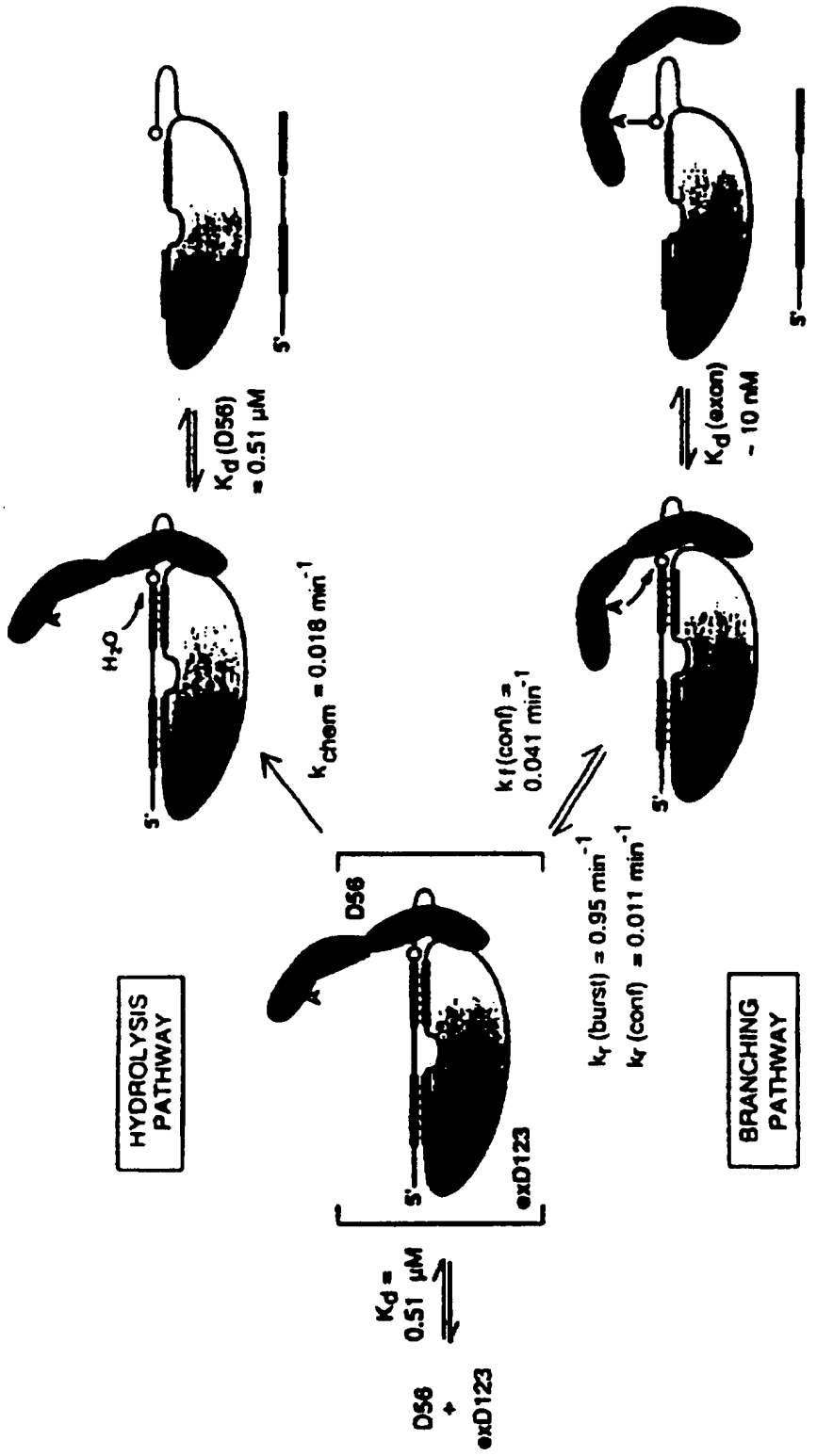

D56
•xD123
5'
$K_d$ = 0.51 µM
D56 + •xD123
HYDROLYSIS PATHWAY
$H_2O$
$k_{chem}$ = 0.018 $min^{-1}$
$k_f$ (burst) = 0.95 $min^{-1}$
$k_f$ (conf) = 0.011 $min^{-1}$
$k_f$(conf) = 0.041 $min^{-1}$
$K_d$(D56) = 0.51 µM
$K_d$(exon) - 10 nM
BRANCHING PATHWAY

FIGURE 15

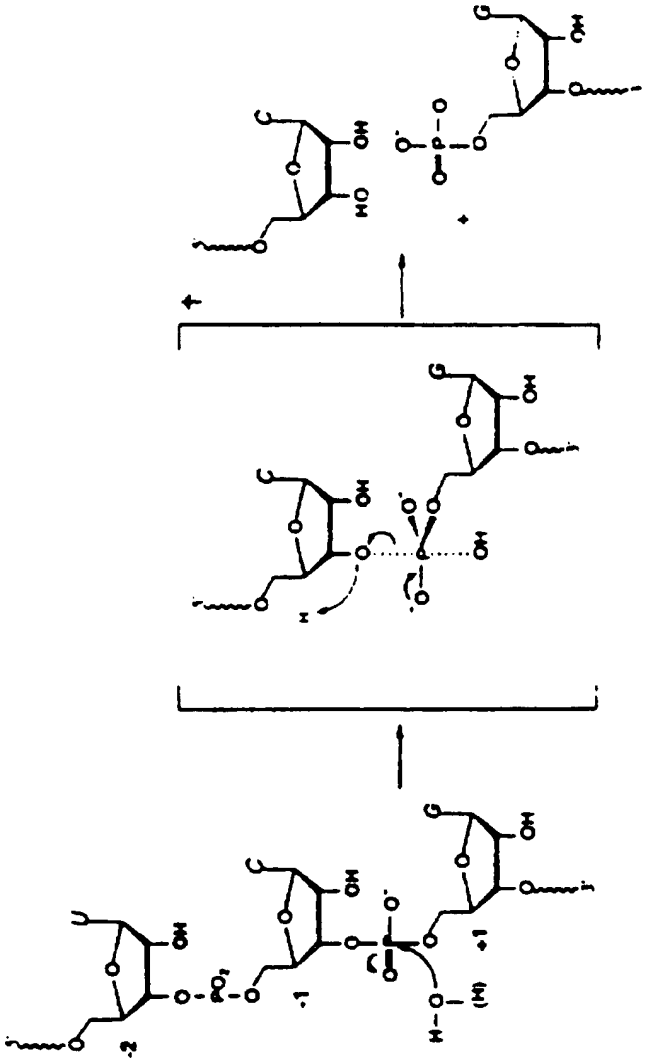


FIGURE 16A

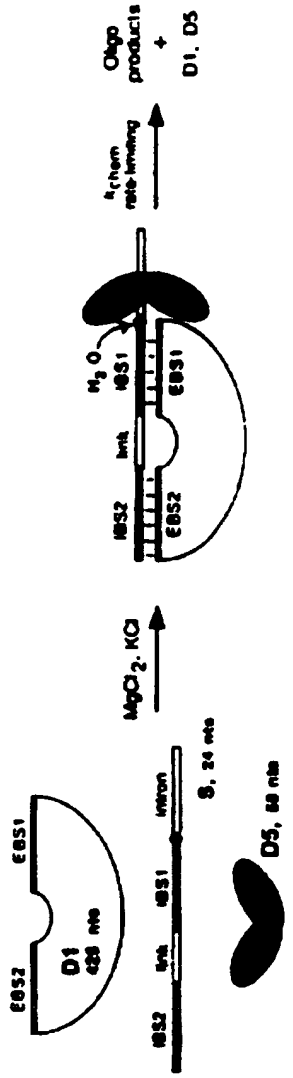
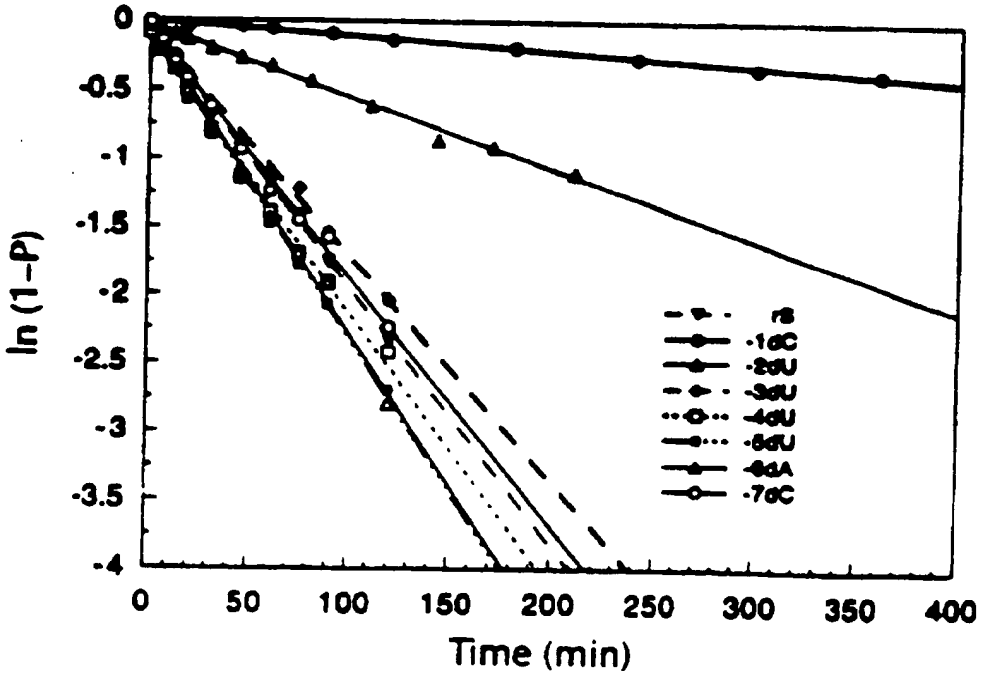


FIGURE 16B

FIGURE 17


0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
ln (1-P)
0
50
100
150
200
250
300
350
400
Time (min)
rB
-1dC
-2dU
-3dU
-4dU
-5dU
-6dA
-7dC

## FIGURE 18

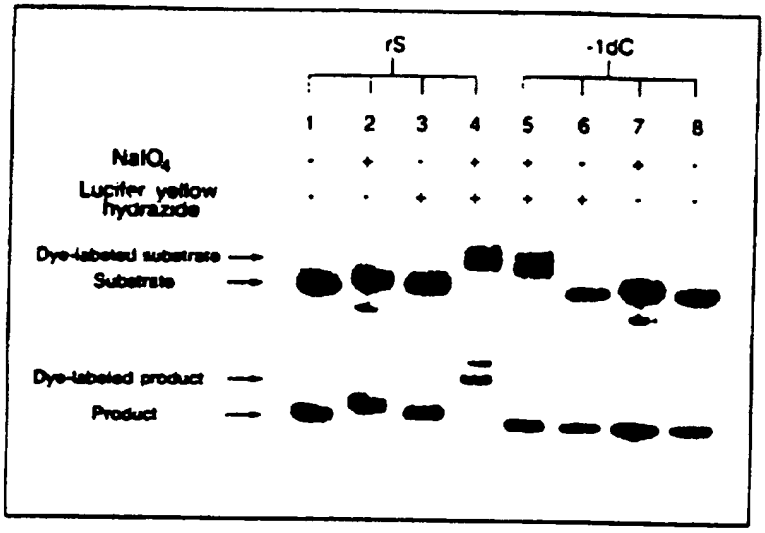

rS
-1dC
1 2 3 4 5 6 7 8
NaIO4
Lucifer yellow hydrazide
Dye-labeled substrate
Substrate
Dye-labeled product
Product

FIGURE 19A

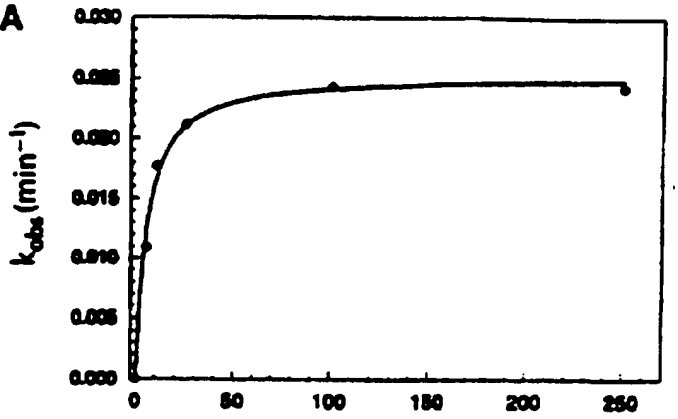



FIGURE 19B

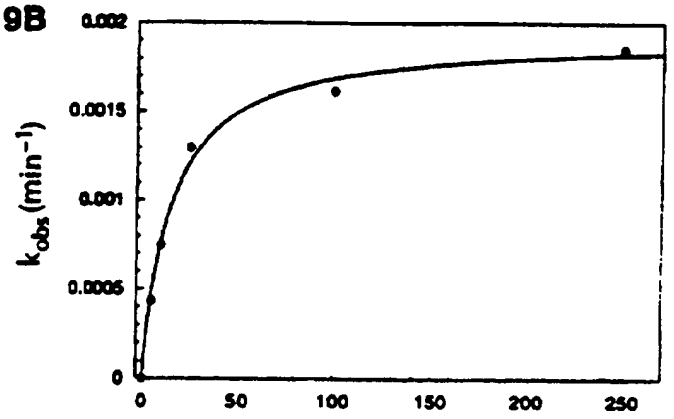



FIGURE 19C

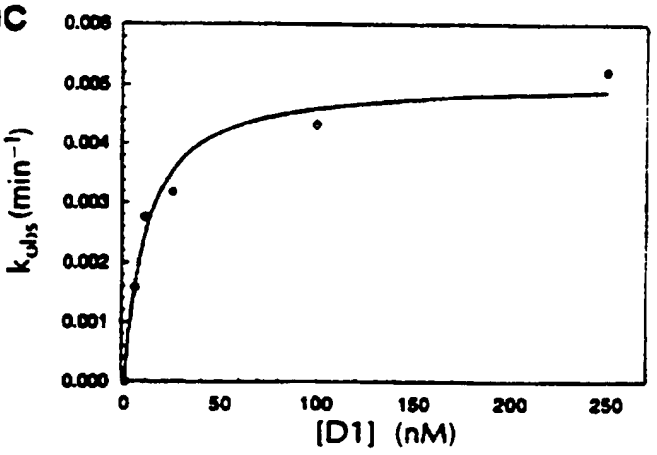


FIGURE 20

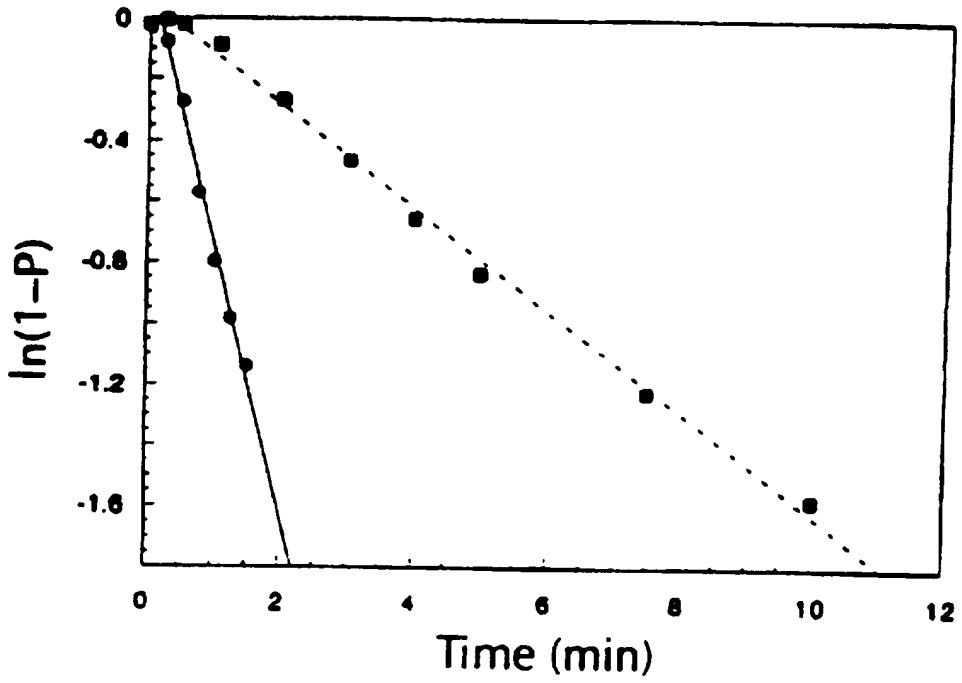


FIGURE 21

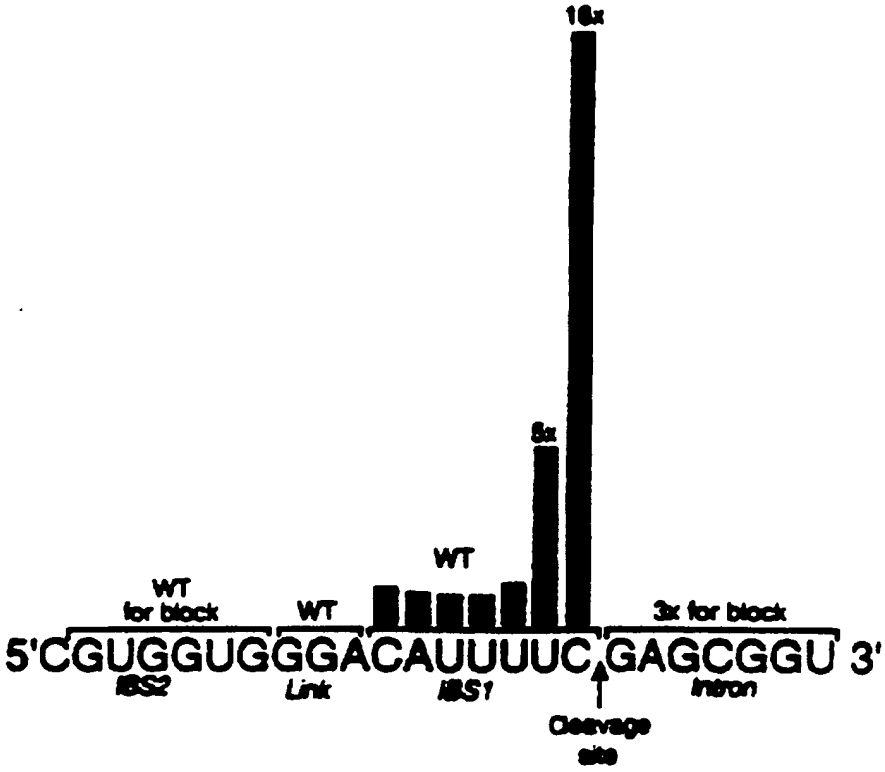

16x
8x
WT
WT
for block
WT
3x for block
5'CGUGGUGGGACAUUUUCGAGCGGU 3'
IBS2
Link
IBS1
Intron
Cleavage
site

FIGURE 22A

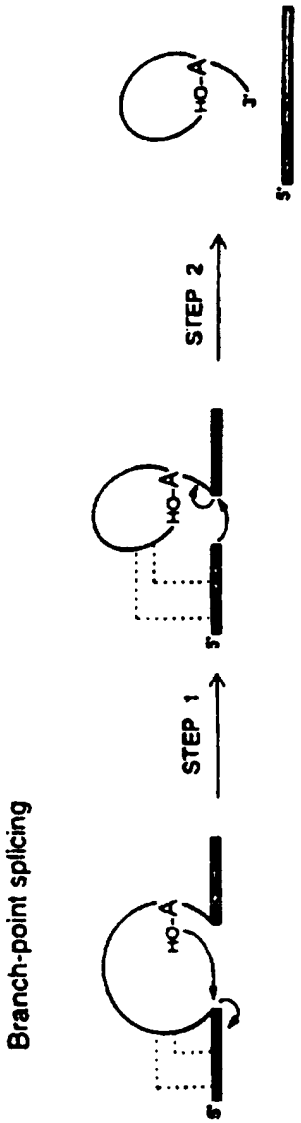


FIGURE 22B

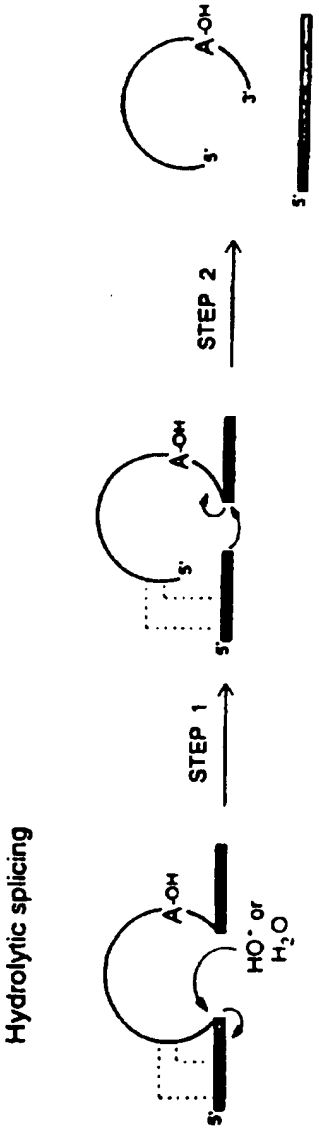


FIGURE 22C

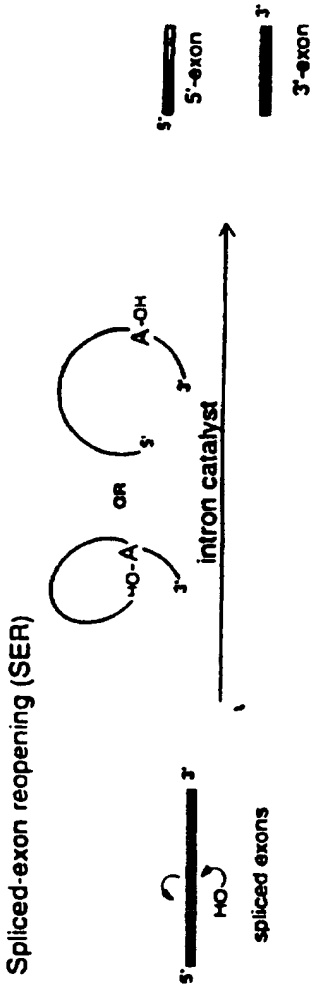

Spliced-exon reopening (SER)
spliced exons
HO
HO-A
OR
A-OH
intron catalyst
5'-exon
3'-exon

FIGURE 23A

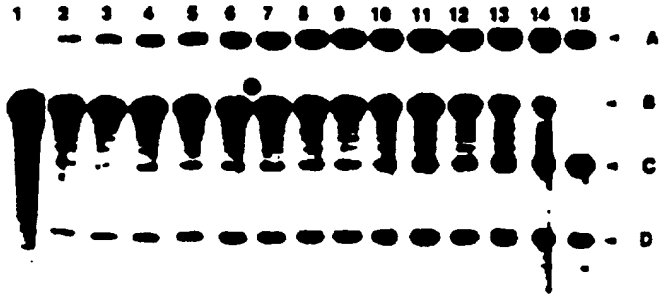



FIGURE 23B

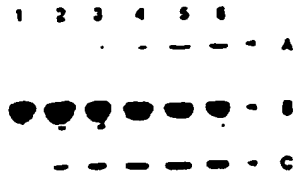



FIGURE 23C

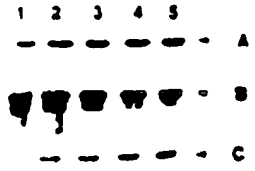
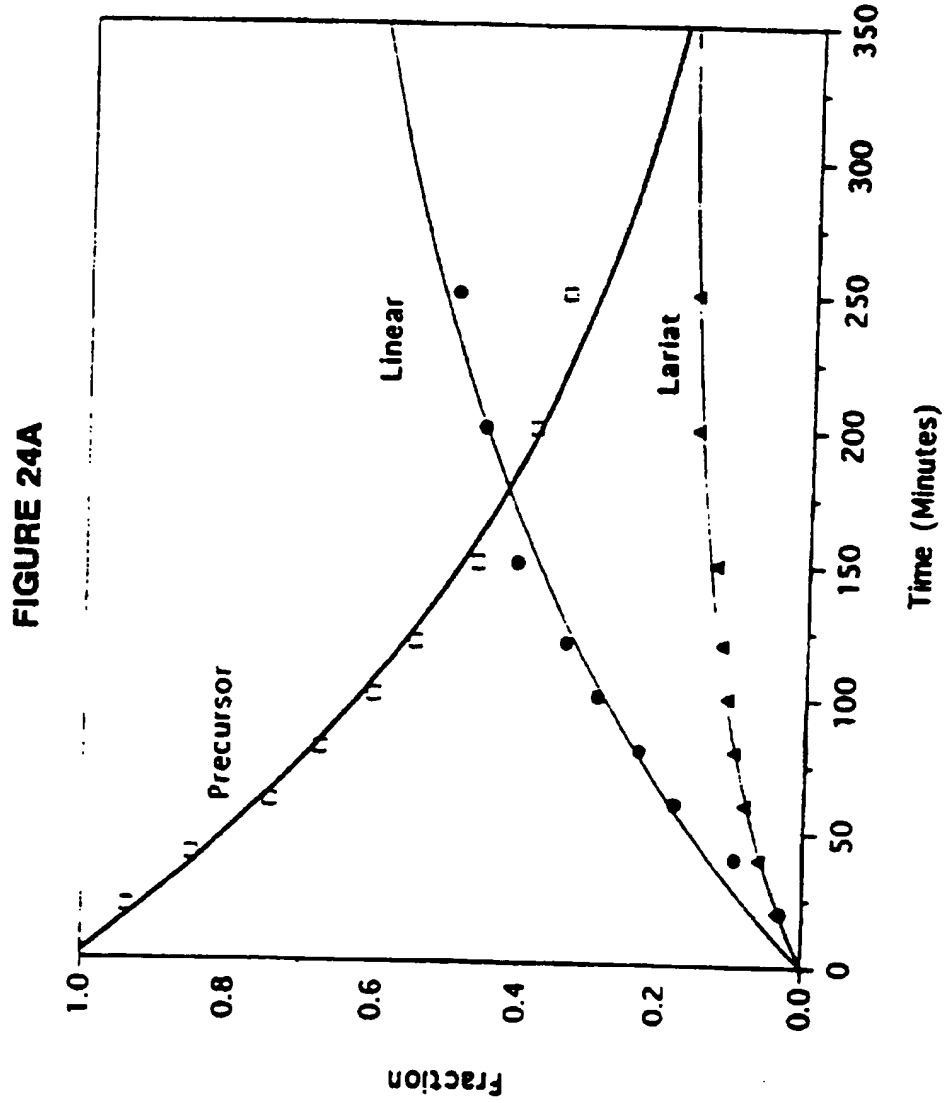


FIGURE 24A


Precursor
Linear
Lariat
1.0
0.8
0.6
0.4
0.2
0.0
Fraction
0
50
100
150
200
250
300
350
Time (Minutes)

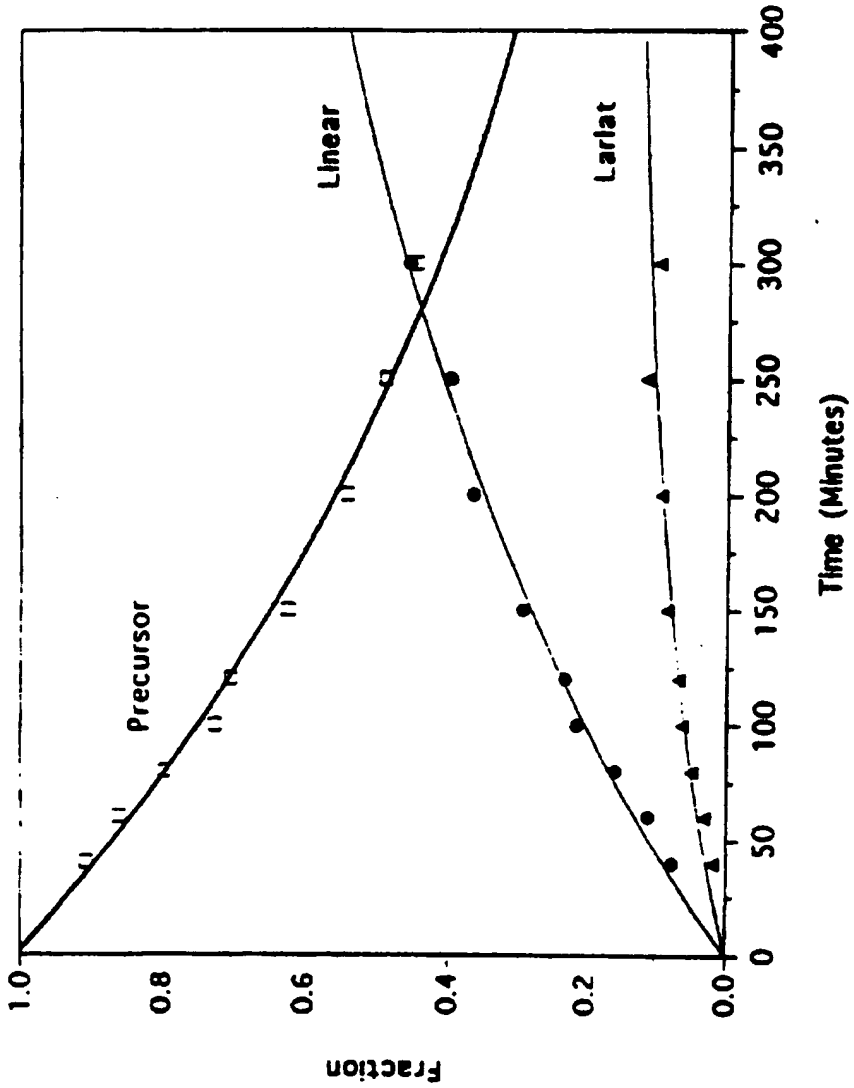

Linear
Lariat
Precursor
Fraction
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
350
400
Time (Minutes)

FIGURE 24B

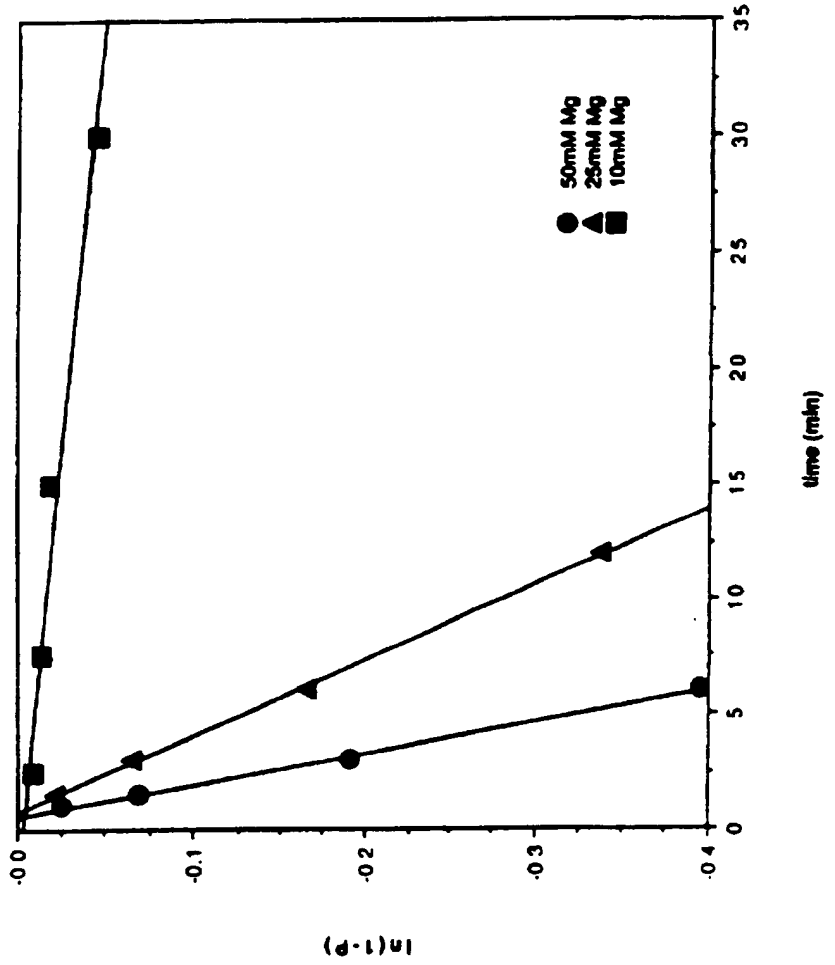

D1,3 enzyme
50mM Mg
25mM Mg
10mM Mg
time (min)
ln(1-P)
0
5
10
15
20
25
30
35
-0.0
-0.1
-0.2
-0.3
-0.4

FIGURE 24C

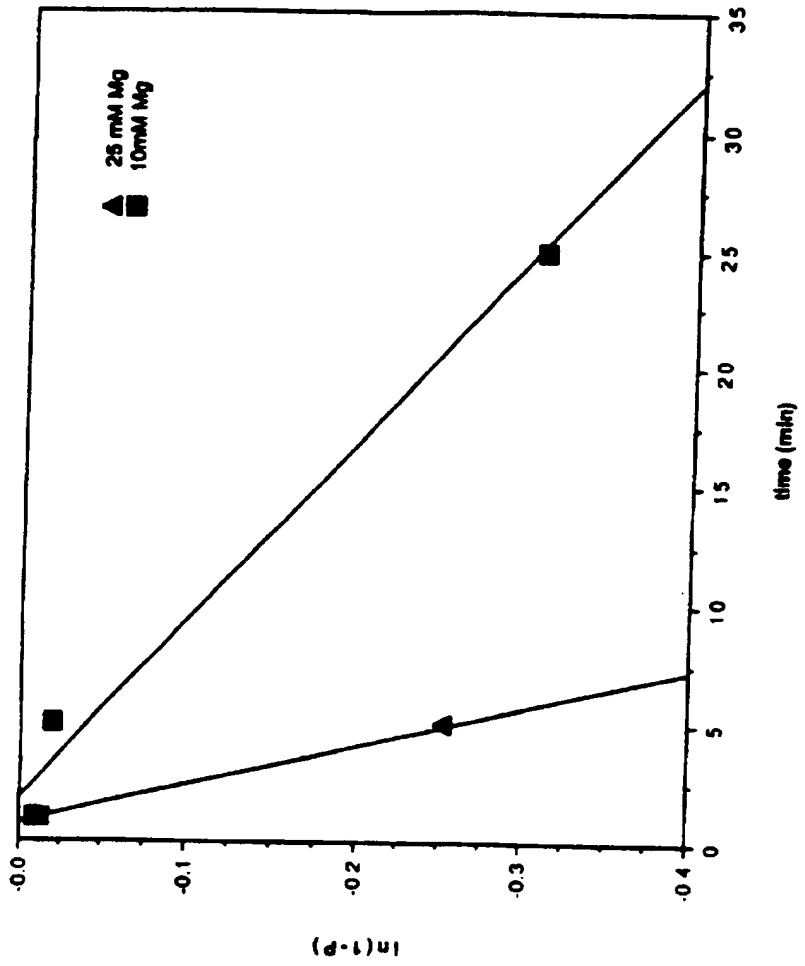

D123 enzyme
25 mM Mg
10mM Mg
ln(1-p)
time (min)
-0.0
-0.1
-0.2
-0.3
-0.4
0
5
10
15
20
25
30
35

FIGURE 24D

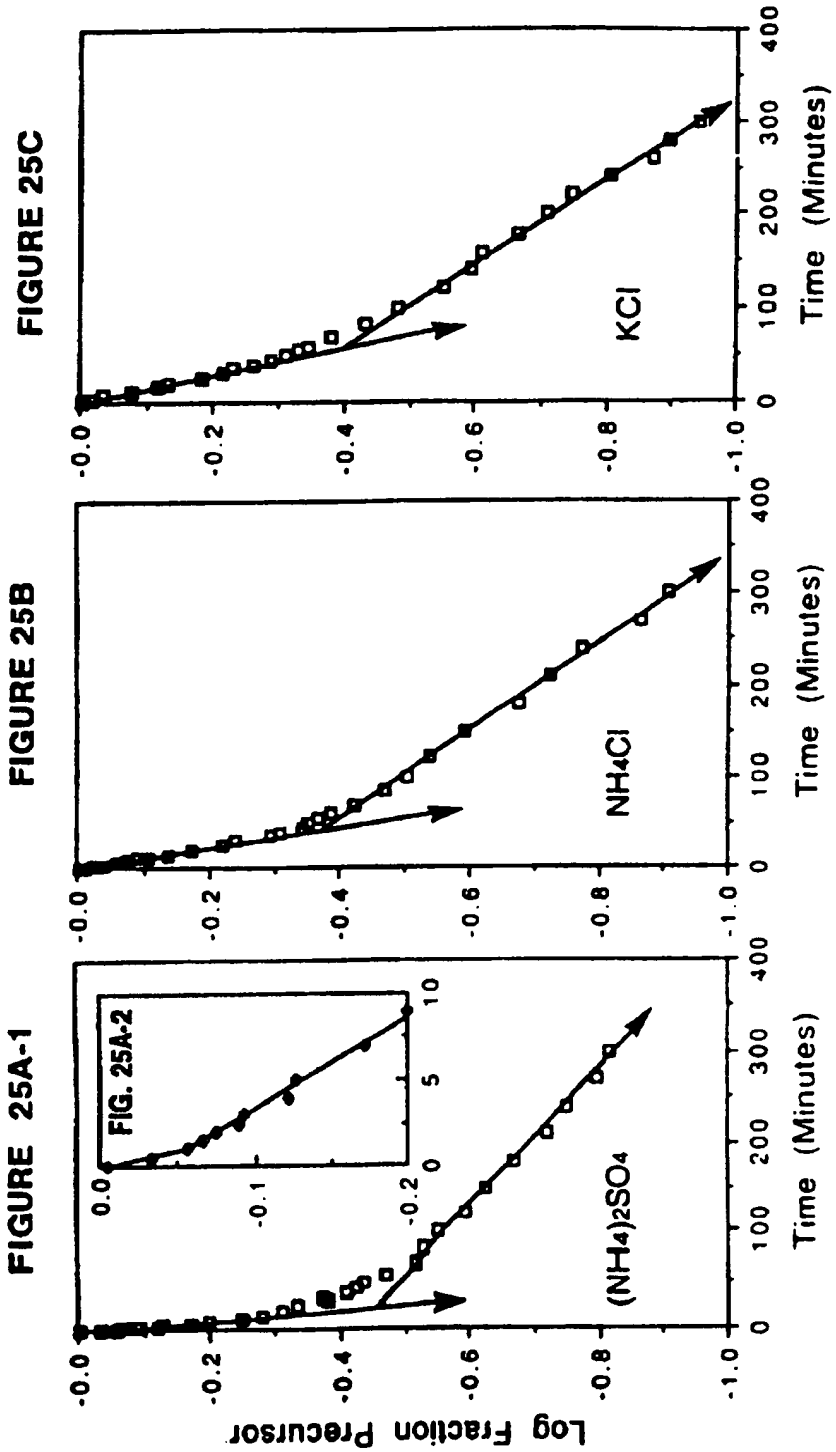


FIGURE 25A-1
FIGURE 25B
FIGURE 25C
Log Fraction Precursor
Time (Minutes)
(NH4)2SO4
NH4Cl
KCl
FIG. 25A-2
0.0
-0.1
-0.2
0
5
10
-0.0
-0.2
-0.4
-0.6
-0.8
-1.0
0
100
200
300
400

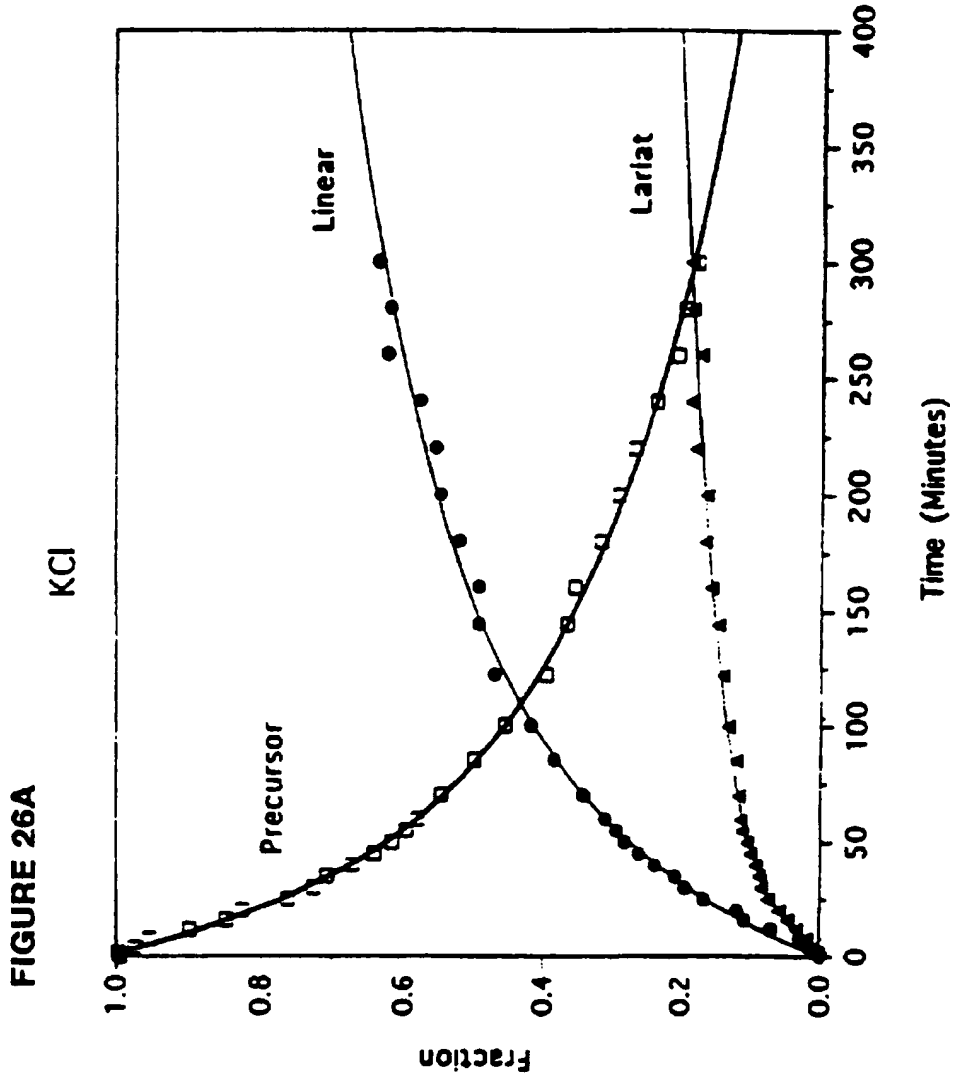

KCl
Linear
Lariat
Precursor
Fraction
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
350
400
Time (Minutes)

FIGURE 26A

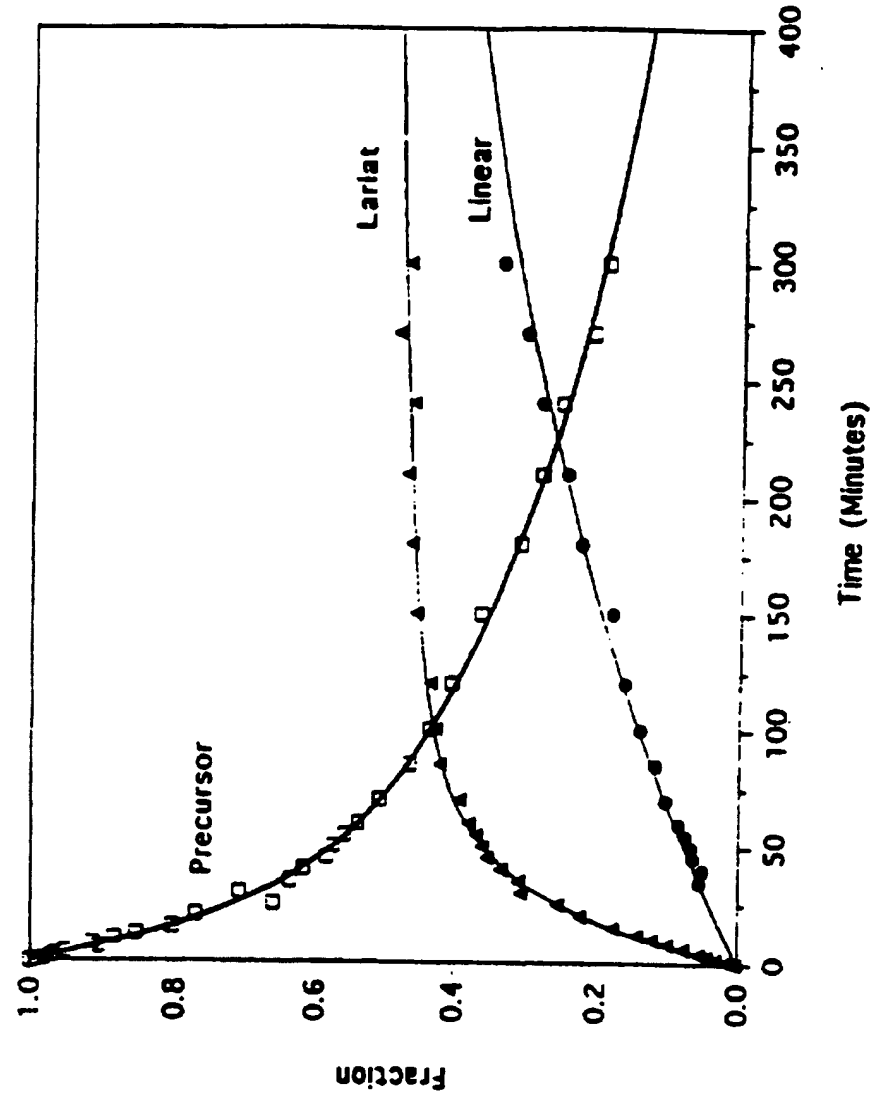

NH4Cl
Precursor
Lariat
Linear
Fraction
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
350
400
Time (Minutes)

FIGURE 26B

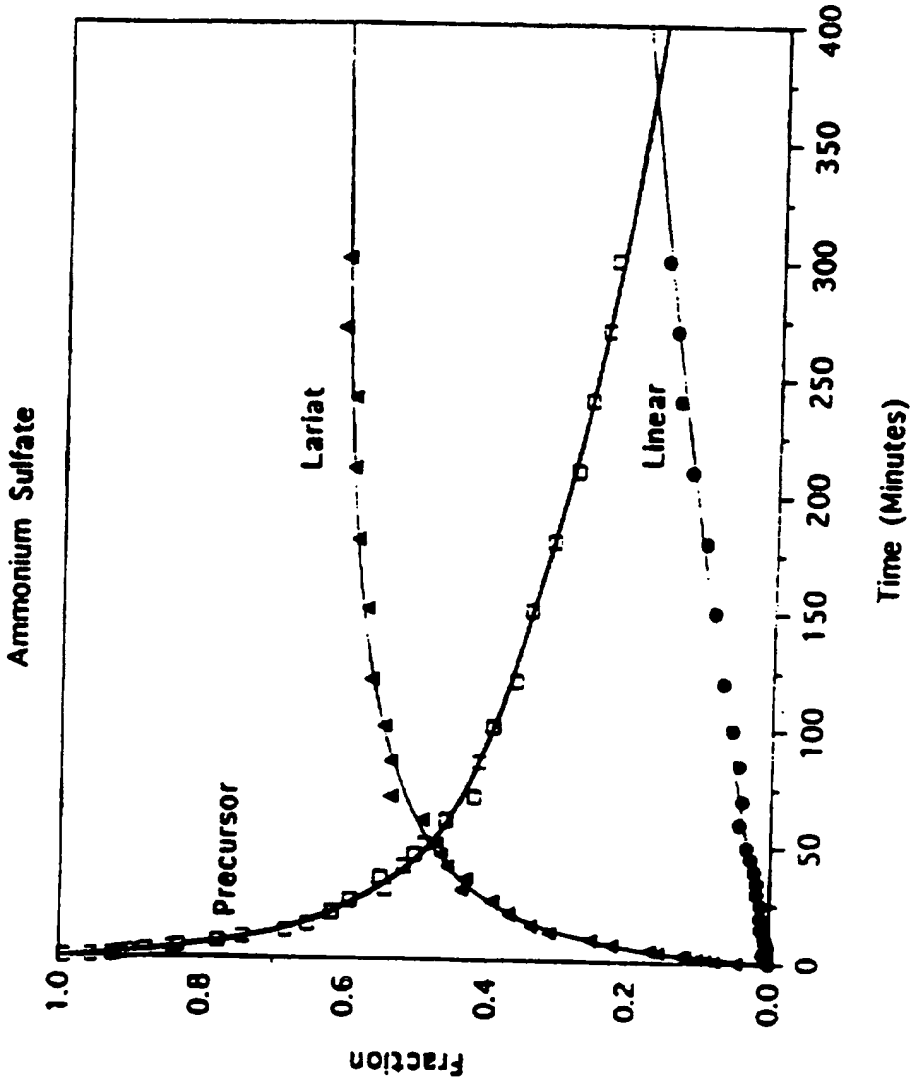
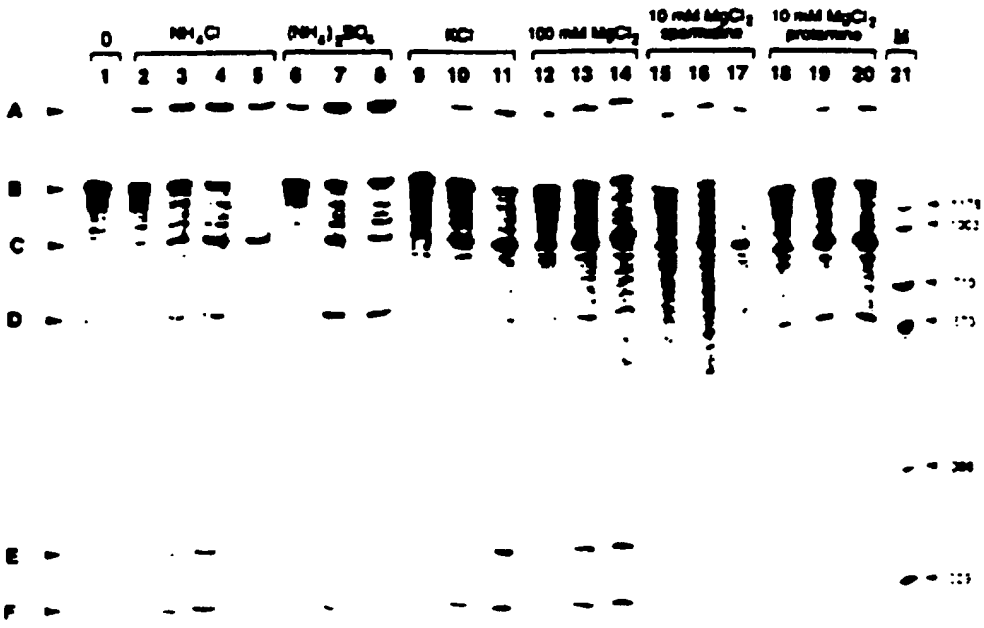

Ammonium Sulfate
Lariat
Linear
Precursor
Fraction
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
350
400
Time (Minutes)

FIGURE 26C

FIGURE 27


0
$NH_4Cl$
$(NH_4)_2SO_4$
KCl
100 mM $MgCl_2$
10 mM $MgCl_2$ spermidine
10 mM $MgCl_2$ putrescine
M
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
A
B
C
D
E
F

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01337

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :A01N 43/04, 63/00; A61K 31/70, 48/00; C12N 1/00, 5/00, 5/06, 5/10, 15/00, 15/09, 15/11
US CL :424/93.2, 93.21; 435/69.1, 172.3, 240.2, 240.21, 320.1; 536/23.1, 24.1; 800/2

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/93.2, 93.21; 435/69.1, 172.3, 240.2, 240.21, 320.1; 536/23.1, 24.1; 800/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Databases: APS, Biosis, Chemical Abstracts, Medline, Embase
Search Terms: ribozym?; intron#; catalyt?; rna; hybridiz?; domain? transesteri?; growth factor?; oncogen?; translocat?; vira?; plant?; vect?; transgen?

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| Y | BIOCHEMISTRY, Volume 29, issued 1990, H. Yanagawa et al., "A novel minimum ribozyme with oxidoreduction activity", pages 10585-10589, see entire document. | 1-57 |
| Y | PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE, USA, Volume 91, issued March 1994, S. Efrat et al., "Ribozyme-mediated attenuation of pancreatic beta-cell glucokinase expression in transgenic mice results in impaired glucose-induced insulin secretion", pages 2051-2055, see entire document. | 1-57 |

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 07 MAY 1996 | 15 MAY 1996 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>BRIAN R. STANTON<br>Telephone No. (703) 308-0196 |

Form PCT/ISA/210 (second sheet)(July 1992)*